Fakultät für Chemie

Fachgebiet Industrielle Biokatalyse

Genetic Engineering of the Oleaginous Yeast *Trichosporon oleaginosus* and the Bacteria *Escherichia Coli* Aimed at the Production of High Value Lipids and Bioactive Diterpenes

Christian Görner

Vollständiger Abdruck der von der Fakultät für Chemie der Technischen Universität München zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften (Dr. rer. nat.) genehmigten Dissertation.

Vorsitzender:	Univ.-Prof. Dr. rer. nat. Tom Nilges
Prüfer der Dissertation	1. Univ.-Prof. Dr. rer. nat. Thomas Brück
	2. apl. Univ.-Prof. Dr.rer.nat Wolfgang Eisenreich
	3. Jun.-Prof. Dr. rer. nat. Robert Kourist (Ruhr Universität Bochum)

Die Dissertation wurde am 24.11.2015 bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie am 04.02.2016 angenommen.

Eidesstattliche Erklärung

Hiermit versichere ich, dass ich die vorliegende Dissertation selbstständig verfasst, sowie die Ausführungen und Gedanken, welche anderen Schriften sinngemäß oder wörtlich entnommen wurden, sowie weitere Quellen und Hilfsmittel kenntlich gemacht habe. Die vorliegende Arbeit wurde bisher weder in gleicher noch ähnlicher Form einer anderen Prüfungsbehörde vorgelegt oder anderweitig veröffentlicht.

München, den _____

Acknowledgements

I would like to thank my supervisor Prof. Dr. Thomas Brück for his constant support, critical review and his open mind for new ideas. Without the liberty I was entitled during my PhD studies, I would not have achieved my research goals and evolved my research skills in so many interdisciplinary fields. Furthermore, I am grateful for the opportunity provided by him to supervise several Bachelor's and Master's thesis. Providing guidance to students did not only push my research into the right direction, but it also helped me with my personal development.

I would like to thank Patrick Schrepfer for his support in bioinformatics, the helpful advices he gave me, for this critical review of data, for the successful collaboration and the great time I had with him in and out of the laboratory.

I would like to thank Veronika Redai for all the help in laboratory, her help with my research, the very useful feedback and the great time we had in and out of the laboratory, especially on our business trip to Prague.

Many thanks go to Martina Haack for her help with analytics, Daniel Garbe for taking over work on the ChiBio project reports and other organizational tasks, which allowed me to completely focus on my research.

I would like to thank Ronja Janke and Dr. Bernhard Loll (Freie Universität Berlin) for their contributions to our successful collaboration.

Furthermore, I would like to thank Prof. Dr. Eisenreich for this great support regarding NMR structure elucidation.

Special appreciation goes to my colleagues/friends Ina Häuslein, Moritz Schreiber, Felix Bracharz, Max Hirte, Jan Lorenzen, Johannes Schmidt, Tom Schuffenhauer, Mahmoud Masri, Matthias Glemser, Gülnaz Celik and Dr. Farah Qoura for always being there to solve problems and drinking beer.

Last but not least I would like to thank my family for all the support I have received during my PhD studies.

Content

Summary

Part I: Production of High Value Lipids by *Trichosporon oleaginosus*

The oleaginous yeast *Trichosporon oleaginosus* ATCC 20509 can accumulate up to 70% triglycerides as dry cell weight when cultivated on agricultural or food waste streams.[1,2,3] The production of high value non-native lipids from waste streams in bio-refineries by *T. oleaginosus* is a sustainable and economically attractive alternative compared to plant and fish oil resources. However, up to now the lack of genetic accessibility and a heterologous gene expression system have hampered engineering approaches of this yeast strain. In this study the first transformation protocol for *T. oleaginosus* based on *Agrobacterium tumefaciens* was established. Strong heterologous gene expression of a YFP reporter protein was achieved using the constitutive promotor from the endogenous glyceraldehyde-3-phosphate dehydrogenase gene.

A proof of concept is presented, which established *T. oleaginosus* as a flexible production platform for tailor made lipids, derived from various monomeric sugars that are major constituents of waste biomass streams. Based on the newly developed protocol, several yeast strains with non-native fatty acid profiles were created by heterologous expression of a series of fatty acid modifying enzymes. Compared to the wild type, recombinant yeast strains showed an increase of α-linolenic acid production from 2.8% up to 21% with respect to the total cellular fatty acid content (TFA). Furthermore, the non-native very long chain unsaturated fatty acids eicosatrienoic and eicosadienoic acid were produced up to 16% (TFA) and 9% (TFA), respectively. These fatty acids are important metabolic intermediates towards the biosynthesis of the pharmaceutically relevant lipid constituents eicosapentaenoic acid and docosahexaenoic acid from marine resources.[4,5] Additionally, *T. oleaginosus* was genetically engineered to produce the high value non-native (*E*-10, *Z*-12) conjugated linoleic acid. This health promoting fatty acid[6] was generated up to 2.6% (TFA)

Part II: Targeted Engineering of Cyclooctat-9-en-7-ol Synthase CotB2

Diterpenes belong to the terpene family and are characterized by a macrocycle encompassing 20 carbon atoms.[7,8] Diterpenes have diverse biological effects such as antimicrobial, antifungal, anti-inflammatory, and cytostatic activity.[9] The chemical synthesis of diterpene macrocycles is challenging and often achieves low yields, the formation of unfavourable isomers and toxic waste streams.[8] By contrast, in nature the formation of complex diterpene macrocycles such as

cyclooctat-9-en-7-ol is generated by diterpene synthases in a highly selective single step transformation using geranylgeranyldiphosphate (GGPP) as a substrate.[10]

The second part of this thesis focused on the engineering of the bacterial cyclooctat-9-en-7-ol diterpene synthases (CotB2) for the production of non-native new carbon macrocycles. The combination of *in silico* modelling and crystallographic studies was used for a targeted engineering approach of CotB2, which led to the production of several non-native diterpene macrocycles. Structure elucidation by GC-MS and NMR resulted in the identification of the two non-natural fusicoccane macrocycle cyclooctat-7-en-3-ol and cyclooctat-1,7-diene as well as the two non-native natural products (1R,3E,7E,11S,12S)-3,7,18-dolabellatriene and R-cembrene A. The formation of the natural product (1R,3E,7E,11S,12S)-3,7,18-dolabellatriene by an engineered CotB2 synthase is the first biotechnological route to this diterpene with antibiotic activity against multidrug-resistant *Staphylococcus aureus*, formally isolated from the brown algae *Dilophus spiralis*.[11] The observed product pattern of the engineered CotB2 synthase was analyzed by *in silico* modelling of the crystallographic results, which gave insights into the structure –function relationship of this diterpene synthase.

Part III: *Escherichia coli* based Production of Cyclooctatin

The third part of this thesis investigates methodologies for the production of multi-functionalized bacterial diterpenoids. *De novo* recombinant production of the multi-functionalized diterpenoids is challenging due to the lack of an efficient redox system that reconstitutes activity of the class I P450 monooxygenases.[12] In this study, the *E.coli* based biotechnological production of cyclooctatin, a potent inhibitor of human lysophospholipase[13] was accomplished. Production of the tri-hydroxylated diterpene cyclooctatin required the reconstitution of the specific hydroxylases CotB3/CotB4, which was accomplished by a newly discovered reductase/ferredoxin system from *S. afghaniensis* (AfR•Afx). The AfR•Afx redox system allowed the efficient production of cyclooctatin with a 43 fold increased yield compared to the native producer. It was demonstrated that the AfR•Afx system is superior in activating the class I P450 hydroxylases CotB3/CotB4 compared to the well-known *Pseudomonas putida* derived PdR•Pdx reference model. The molecular basis for these activity differences was investigated with regard to finding a general approach to increase the redox efficiency between class I P450s and non-native redox systems. Therefore, a case study on the PdR•Pdx reference system was conducted. It was demonstrated that specific redox system engineering can boost and harmonize the catalytic efficiency of class I hydroxylase enzyme cascades.

Furthermore, the efficient reconstitution of the P450 hydroxylases CotB3/CotB4 allowed the evaluation of the substrate promiscuity by screening a series of diterpene macrocycles. This approach led to the discovery of the first biotechnological route for the production of sinularcasbane D, a functionalized diterpenoid from the soft coral *Sinularia sp.*[14]

Zusammenfassung

Teil I: Herstellung von hochwertigen Lipiden in Trichosporon oleaginosus ATCC 20509

Die Öl-Hefe *Trichosporon oleaginosus* ATCC 20509 zeichnet sich durch die Fähigkeit aus, große Mengen an Lipiden aus biogenen Reststoffen der Agrar- und Lebensmittelindustrie anzureichern, die bis zu 70% der Zellbiotrockenmasse ausmachen können.[1,2,3] Der fermentative Einsatz von *T. oleaginosus* in einer Bioraffinerie zur Produktion von hochwertigen artifiziellen Lipiden aus biogenen Reststoffen stellt daher eine ökonomisch sinnvolle und nachhaltige Alternative gegenüber der klassischen Produktion von Pflanzen- und Fischölen dar. Aufgrund des Mangels an geeigneten Transformations- und heterologen Expressionssystemen waren gentechnische Arbeiten an der Öl-Hefe bisher nur stark eingeschränkt durchführbar. Im Rahmen dieser Arbeit wurde dazu ein Transformationsprotokoll für *T. oleaginosus* basierend auf dem Bakterium *Agrobacterium tumefaciens* entwickelt. Darüber hinaus wurde gezeigt, dass die heterologe Genexpression des YFP Reporterproteins mit Hilfe des starken konstitutiven Promotors aus dem Glycerinaldehyd-3-phosphat-Dehydrogenase Gens möglich ist.

In einer Machbarkeitsstudie wurde gezeigt, dass *T. Oleaginosus* als flexible Plattform zur Herstellung von maßgeschneiderten Lipiden aus monomeren Zuckern von biogenen Abfallstoffströmen verwendet werden kann. Hierzu wurden Hefestämme hergestellt, die mit Hilfe von Fettsäure modifizierenden Enzymen eine Vielzahl hochwertiger nicht nativer Fettsäuren produzierten. So konnte gezeigt werden, dass in einer gentechnisch veränderten Hefe der Anteil der α-Linolensäure an der Gesamtmenge der Fettsäuren von 2.8% auf 21% gestiegen war. In einer weiteren gentechnisch veränderten Hefe wurden die nicht nativen langkettigen ungesättigten Fettsäuren Eicosatriensäure mit 16% und Eicosadiensäure mit 9% am Anteil der Gesamtmenge der Fettsäuren erfolgreich produziert. Diese Fettsäuren sind wichtige metabolische Zwischenstufen bei der Biosynthese der pharmakologisch relevanten Fettsäuren Eicosapentaensäure (EPA) und Docosahexaensäure (DHA) aus marinem Ursprung.[4,5] Darüber hinaus wurde bewiesen, dass *T. oleaginosus* zur Produktion der nicht nativen (*E*-10, *Z*-12) konjungierten Linolsäure (2.6% Fettsäureanteil) geeignet ist. Letzterer werden zahlreiche gesundheitsfördernde Effekte zugesprochen.[6]

Teil II: Zielgerichtetes Engineering der Diterpene Synthase CotB2

Diterpene gehören zur Naturstoffklasse der Terpene und beschreiben Makrozyklen mit 20 Kohlenstoffatomen.[7,8] Vertreter der Diterpene können antibiotisch, antifungal, entzündungshemmend und zytostatisch wirken und sind deshalb von großem pharmakologischem Interesse.[9] Die klassische chemische Naturstoffsynthese von Diterpenen ist oft sehr anspruchsvoll, auch liefert sie eine geringe Ausbeute bei gleichzeitiger Produktion großer Abfallmengen.[8] Im Gegensatz dazu werden biochemisch Diterpen-Makrozyklen in einem einzigen Syntheseschritt aus dem Vorläufer-Molekül Geranylgeranyldiphosphat (GGPP) hoch effizient und stereoselektiv aufgebaut.[10]

Im zweiten Teil dieser Arbeit wurden Enzym-Mutanten der CotB2 Diterpen-Synthase zur Synthese von neuen nicht natürlichen Diterpen-Makrozyklen hergestellt. Mit Hilfe von bioinformatischen Modellierungen und kristallografischen Untersuchungen wurden dazu Bereiche innerhalb der CotB2 Synthase ausgewählt und gezielt verändert, was zur Entdeckung von neuen nicht nativen Diterpenen führte. Durch Strukturaufklärung mittels GC-MS und NMR konnten die zwei nicht natürlichen Fusicoccane Cyclooctat-7-en-3-ol und Cyclooctat-1,7-dien., sowie die Naturstoffe (1R,3E,7E,11S,12S)-3,7,18-Dolabellatrien und R-cembrene identifiziert werden. Die Herstellung von (1R,3E,7E,11S,12S)-3,7,18-Dolabellatriene mit Hilfe einer angepassten CotB2 Synthase, ist die erste biotechnologische Route dieses Naturstoffs, der zuvor in der Braunalge *Dilophus spiralis* entdeckt wurde und einen anti-mikrobiellen Effekt gegen multi-resistente *Staphylococcus aureus* Stämme besitzt.[11] Die beobachteten Syntheseigenschaften der veränderten CotB2 Synthase wurden durch computergestützte Simulation untersucht. Letzteres führte zu Einblicken in die Struktur-Funktions-Beziehung.

Teil III: Herstellung von Cyclooctatin im Bakterium *Escherichia Coli*

Der dritte Teil dieser Arbeit untersucht Methoden zur bakteriellen Herstellung von multifunktionellen Diterpenoiden. Die Herstellung von mehrfach hydroxylierten Diterpenen ist anspruchsvoll, da hierzu häufig mehrere Klasse I Cytochrome P450 Monooxygenasen funktionell in *E.coli* integriert werden müssen.[12] Oftmals scheitert dieser Ansatz am Mangel eines passenden Reduktase/Ferredoxin Systems, welches die Aktivität der P450 Hydroxylase Enzyme gewährleistet.[12] In dieser Studie wurde die biotechnologische Produktion von Cyclooctatin, ein Inhibitor der humanen Lysophospholipase[13], in *E.coli* realisiert. Die Herstellung des dreifach hydroxilierten Diterpens Cyclooctatin erforderte die Rekonstitution der spezifischen Hydroxylasen CotB3/CotB4, die durch ein neu entdecktes Reduktase/Ferredoxin-

System aus *S. Afghaniensis* (AfR•Afx) ermöglicht wurde. Das AfR•Afx Redoxsystem ermöglichte die effiziente Herstellung von Cyclooctatin mit 43-fach höherer Ausbeute gegenüber der nativen Produktion. Ferner wurde das neu entdeckte AfR•Afx Reduktasesystem mit dem Literatur bekannten PdR•Pdx System aus dem Bacterium *Pseudomonas putida* verglichen. Es zeigte sich, dass das AfR•Afx System die P450 Hydroxylasen CotB3/CotB4 signifikant besser aktiviert. Die molekularen Ursachen für die beobachteten Aktivitätsunterschiede wurden mit dem Ziel untersucht, eine generelle Vorgehensweise zur Verbesserung der Redoxwirkung zwischen Klasse I P450 Hydroxylasen und nicht-nativen Redoxsystemen zu finden. Zu diesem Anlass wurde eine Fallstudie mit dem PdR•Pdx Referenzsystem durchgeführt. Es konnte gezeigt werden, dass durch spezifische Änderungen die Aktivität der P450 Hydroxylasen erhöht und gleichzeitig die Balance zwischen den P450 Hydroxylasen verbessert werden kann.

Ein weiterer Teil dieser Arbeit widmete sich der Evaluierung der Substrat Promiskuität der P450 Hydroxylasen CotB3/CotB4. Dabei wurde die erste biotechnologische Route für die Produktion des Naturstoffs Sinularcasbane D aus der Weichkoralle *Sinularia sp.* entdeckt.[14]

List of Related Articles

The thesis includes the following related articles:

(1) Görner, C.; Häuslein, I.; Schrepfer, P.; Eisenreich, W.; Brück, T. *ChemCatChem* **2013**, *5*, 3289–3298.

(2) Janke, R.; Görner, C.; Hirte, M.; Brück, T.; Loll, B. *Acta Crystallogr. D. Biol. Crystallogr.* **2014**, *70*, 1528–1537.

(3) Görner, C.; Hirte, M.; Huber, S.; Schrepfer, P.; Brück, T. *Front. Microbiol.* **2015**, *6*, 1115.

(4) Görner, C.; Redai, V.; Bracharz, F.; Schrepfer, P.; Garbe, D.; Brück, T. *Green Chem.* **2016**, *18*, 2037–2046.

(5) Görner, C.; Schrepfer, P.; Redai, V.; Wallrapp, F.; Loll, B.; Eisenreich, W.; Haslbeck, M.; Brück, T. *Microb. Cell Fact.* **2016**, *15*, 86.

List of Abbreviations and Symbols

(*v*/*v*)	Volume per volume	IDI	Isopentyl pyrophosphate isomerase
(*w*/*w*)	Weight per weight	IMAS	Induction medium with acetosyringone
µg	Microgram	IPP	Isopentenyl diphosphate
µL	Microliter	IPTG	Isopropyl-β-D-thiogalactopyranosid
µM	Micromolar	IspD	4-phosphate cytidylyltransferase
^{3}J	Three-bond coupling across single bonds	IspE	4-diphosphocytidyl-2-C-methyl-D-erythritol kinase
^{4}J	Four-bond coupling across single bonds	IspF	2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase
Å	Angstrom	IspG	(E)-4-Hydroxy-3-methyl-but-2-enyl pyrophosphate synthase
AAS	Actoacetyl-CoaA thiolase	L	Liter
AC	Isocitrate aconitase	LA	Linoleic acid
ACL	ATP-citrate lyase	LB	Left boarder
ACP	Acetyl-carrier protein	LiAc	Lithiumacetate
AdR	Bovine adrenodoxin reductase	L-IMAS	Liquid Induction medium with acetosyringon
Adx	Bovine adrenodoxin	LPA	lysophospatidic acid
AfR	*Streptomyces afghaniensis* reductase	m	Multiplet
Afx	*Streptomyces afghaniensis* ferredoxin	ME	Malate enzyme
ALA	α-linolenic acid	MEcPP	2-C-methyl-D-erythritol 2,4-cyclopyrophosphate
AMP	Adenosine monophosphate	MEP	3-C-methylerythritol-4-phopshate
AMPD	AMP deaminase	MEV	Mevalonate
ATCC	American Type Culture Collection	mg	Milligram
ATMT	*Agrobacterium* tumefaciens mediated transformation	MHz	Megahertz
ATP	Adenosine triphosphate	min	Minutes
aU	Arbitrary units	MK	Mevalonate kinase
bp	Basepair	mL	Milliliter
BSTFA	N,O-Bistrifluoroacetamide	mM	Millimolar
CD	Circular dichroism	MNNG	N-methyl-N'-nitro-N'-nitrosoguanidine
cDNA	Complementary DNA	MS	Mass spectrometry
CDP-ME	4-diphosphocytidyl-2-C-methylerythritol	MTPA	α-Methoxy-α-trifluoromethylphenylacetic acid
CDP-MEP	4-diphosphocytidyl-2-C-methyl-D-erythritol 2-phosphate	MUSTANG	A multiple structural alignment algorithm
CLA	Conjugated linoleic acids	NacGlc	N-Acetylglucosamine
CoA	Coenzyme A	NAD	Nicotinamide adenine dinucleotide
COSY	Correlated Spectroscopy	NADP	Nicotinamide adenine dinucleotide phosphate
CPP	Copalyl diphosphate	NCBI	National Center for Biotechnology Information
CrTE	Geranylgeranyldiphosphate synthase	ng	Nanogram

d	Doublet	NMR	Nuclear magnetic resonance
DAG	Diacylglycerol acyltransferase	NOE	Nuclear Overhauser effect
DCW	Dry cell weight	NOESY	Nuclear Overhauser effect spectroscopy
DEPT	Distortionless enhancement by polarization transfer.	OriV	Origin of replication V
DGAT	Diacylglycerol acyltransferase	PAGE	Polyacrylamide gel electrophoresis
DHA	Docosahexaenoic acid	PAP	Phosphohydrolase
DMAPP	Dimethylallyl pyrophosphate	PdR	*Pseudomonas putida* reductase
DNA	Deoxyribonucleic acid	Pdx	*Pseudomonas putida* ferredoxin
dNTP	Desoxynukleosidtriphosphate	PEG	Polyethylene glycol
DSMZ	Deutsche Sammlung von Mikroorganismen und Zellkulturen	PHA	Polyhydroxyalkanoate
DTT	Dithiothreitol	PMD	Mevalonate-5-pyrophosphate decarboxylase
DXP	1-dexy-D-xylose-5-phosphate	PMK	Phosphomevalonate kinase
DXP	1-deoxy-D-xylose-5-phosphate	ppm	Parts per million
DXS	1-Deoxy-D-xylulose 5-phosphate synthase	PUFA	Poly unsaturated fatty acid
EDA	Ecosadienoic acid	r.m.s.d	Root-mean-square deviation of atomic positions
EDTA	Ethylenediaminetetraacetic acid	RB	Right boarder
ent-CPP	Ent-copalyl pyrophosphate	rpm	Revolutions per minute
EPA	Eicosapentaenoic acid	s	Singlet
ER	Endoplasmic reticulum	SCO	Single cell oil
ETE	Eicosatrienoic acid	SDS	Sodium dodecyl sulfate
E-value	Expected value	S-IMAS	Solid induction medium with acetosyringon
FAD	Flavin adenine dinucleotide	spFdR	*Spinach* reductase
FAME	Fatty acid methyl ester	spFdx	*Spinach* ferredoxin
FID	Flame ionisation detector	SS	Single stranded DNA
FPP	Farnesylpyrophosphate	t	Triplet
G3P	Glycerol-3-phosphate	T-DNA	Transfer DNA
GAT	Glycerol-3-phosphate acyltranferase	TFA	Total cellular fatty acid content
GC	Gaschromatography	TGA	Triacylglycerol
GGDP	Geranylgeranylpyrophosphate	Ti	Tumor inducing
Glc	Glucose	TMHMM	Membrane protein topology prediction method based on a hidden markov model
GPD	Glyceraldehyde-3-phosphate dehydrogenase	TrfA	Replication initiation protein
GPP	Geranylpyrophosphate	tRNA	Transfer ribonucleinic acid
h	Hour	TrpyC	Tryptophan synthase C
HHpred	HHpred - Homology detection & structure prediction by HMM-HMM comparison	U	Unit
HMBC	Heteronuclear multiple-bond correlation spectroscopy	USER	Uracil-specific excision reagent
HMB-PP	(E)-4-Hydroxy-3-methyl-but-2-enyl pyrophosphate	VLC	Very long chain

HMF	Hydroxymethylfurfural	VLCPUFA	Very long chain poly unsaturated fatty acids
HMGR	3-hydroxy-3-methylglutaryl-CoA reductase	vvm	Gas volume flow per unit of liquid volume per minute
HMGS	3-hydroxy-3-methylglutaryl-CoA synthase	Xyl	Xylose
HMM	Hidden markov model	Yasara	Yet Another Scientific Artificial Reality Application
hph	Resistance B resistance gene	YFP	Yellow fluorescent protein
HSQC	Heteronuclear single-quantum correlation spectroscopy	YPD	Yeast extract peptone dextrose
ICDH	Isocitrate dehydrogenase		

1. General Introduction

1.1 Biocatalysis and the Concept of a Bio-Refinery

As the global population grows to approximately 9-10 billion people by 2050, fulfilling the demands for energy, clean water, and food becomes increasingly challenging.[15,16,17] At the same time the untamed burning of fossil resources intensifies climate change and presents a threat to the global biodiversity.[18] As a result there is an increasing desire to seek sustainable, environmentally friendly, and bio-based processes for the production of energy, fuel, chemical commodities and pharmaceuticals.[19,20,21,22,23] The development of environmentally friendly processes encompasses key parameters such as the reduction of waste, increased atom efficiency, energy efficiency, and avoidance of toxic reagents and hazardous solvents.[24,25,26] These principles are the fundament of the 12 Principles of Green Chemistry[27] and are applied in today's chemical synthesis by the development of less energy requiring and less waste generation technologies.[28] In particular, chemical synthesis based on biocatalytical reaction steps have significantly expanded the portfolio of environmentally friendly processes.[24] Enzyme and whole-cell biocatalytical processes display a variety of advantages compared to traditional chemical processes that makes them a preferred candidate for green chemistry.[24] Biochemical processes are regarded as more energy efficient, less raw material consuming and less waste stream producing than classical chemical processes.[24] Enzymes and microorganisms applied in biochemical productions are water based non-toxic catalysts, in contrast to classical organic solvent based reactions.[24] Microorganisms can produce commodity chemicals, fine chemicals or pharmaceuticals from renewable bio-based substrates such as sugars, lipids and CO_2.[29,30,31,32,33,34,35] These biocatalytical processes take place at ambient temperature, pressure and pH levels, which significantly reduces energy consumption compared to traditional chemical steps.[24] Furthermore, biocatalytical approaches often shorten classical chemical synthesis and are highly selective in terms of regio- and stereoselectivity.[24]

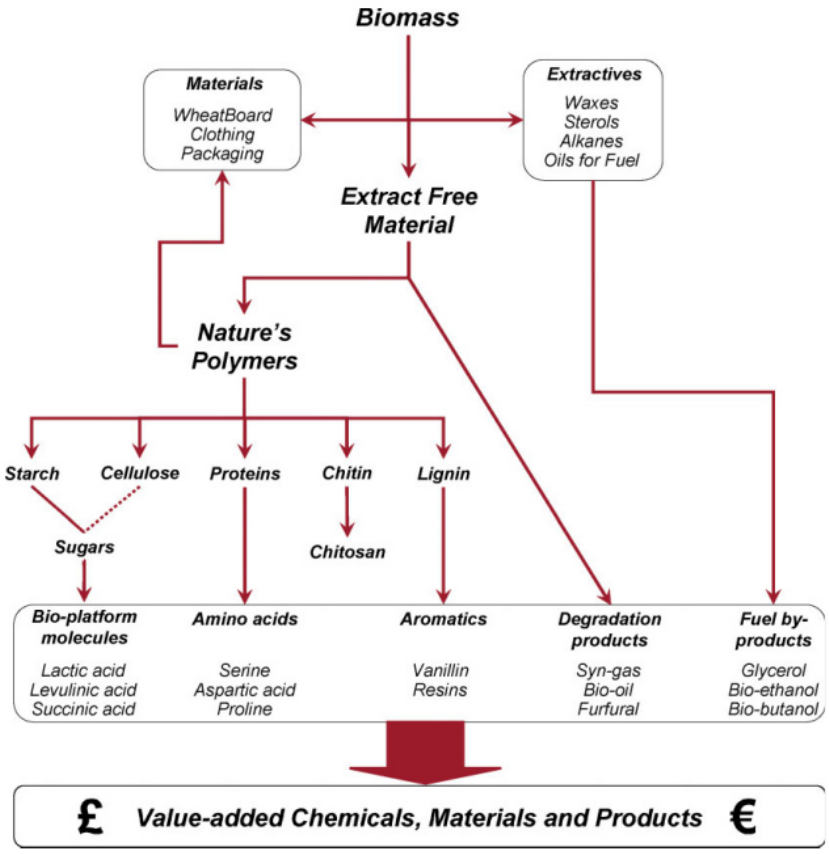


Figure 1.[22] A simplified and idealized bio-refinery for the production of commodity chemicals.

The concept of a bio-refinery comprises the principles of green chemistry and applies them to the catalytic transformation of biomass into its building blocks, which are used for the production of value added chemicals, materials, biofuels and bioenergy.[19,21,22,36] A bio-refinery is an industrial facility or a network of facilities, which uses an extensive range of combined technologies that may contain biocatalytical and chemical routes.[19] In contrast to fossil oil refineries, which produce multiple fuels and chemicals from petroleum, a bio-refinery aims at the full sustainable transformation of biomass into value added products and energy without side products.[19,22] Figure 1 illustrates the processes of an idealized bio-refinery with the focus on commodity chemicals. The biomass is separated into extractives (e.g waxes, sterols, alkanes and oils) and biopolymers (e.g starch, cellulose, proteins, chitin and lignin).[22] Both components are converted into a vast array of value added chemicals.[22] Alternatively, biopolymers can be directly used to manufacture bio based materials (e.g wheatboard) or are remodeled into biopolymers (e.g. polylactic acid, inulin (polyfructose)) for packaging and clothing.[22] Ideally, processing uses chemicals obtained from the extractable fraction of the biomass.[22]

Bio-refineries can be classified according to the biomass feedstock used.[37] First generation feedstocks comprise food crops such as corn, wheat and sugar cane.[37] Today about 80% of the

biofuel production consists of ethanol produced from first-generation feedstocks.[37] This has been criticized due to the competition with the food production.[37] Second generation feedstocks are lignocellulosic materials from energy crops such as poplar and miscanthus as well as waste streams from agriculture, forestry and municipality.[37] Biomass from microalgae is considered as the third generation feedstock and can be grown using CO_2 as substrate from power plants or industry.[37] Although there are currently bio-refineries in commercial operation and demonstration scale that convert first and second generation biomass into multiple products, the full concept of a bio-refinery has not been realized to date.[19,22] An example of a modest bio-refinery is a sugar cane mill that produces few products such as sugar, ethanol and polyhydroxybutyrate as value added products.[19] In the future the development of bio-refineries is expected to expand the variety of feedstocks useable and the range of value added sustainable products generated.[19] Furthermore, maximizing the biomass conversion efficiency will help to reduce the demand for raw materials.[19] The development and integration of new green chemical processes in bio-refineries will make a significant contribution to a sustainable bio-economy.[19]

2. Part I: Production of High Value Lipids by *Trichosporon oleaginosus*

Peer-Reviewed Publication

This chapter is based on the following article: "Genetic engineering and production of modified fatty acids by the non-conventional oleaginous yeast Trichosporon oleaginosus ATCC 20509" Görner, C.; Redai, V.; Bracharz, F.; Schrepfer, P.; Garbe, D.; Brück, T. *Green Chem.* **2016**, *18*, 2037–2046.

Contributions

CG, TB conceived the study, planned and supervised experiments. CG and VR conducted the experiments and analytics. CG performed the data analysis. VR and FB contributed with data analysis. CG, VR, FB, PS, DG and TB prepared the manuscript.

2.1 Introduction

2.1.1 The ChiBio Project: A Chitin Based Bio-Refinery

The fishing industry in the European Union (EU) and worldwide produces an increasing amount of crustacean shell waste that is currently disposed in landfills.[38] This practice represents a significant financial burden and puts human health and the environment at risk.[38] By contrast, shrimp shells from Asia are commercially recycled into the value added product chitosan, which is used in agriculture to stimulate plant growth, for seed coating or in several medical applications.[38,39,40] A commercial application for the European crustacean shell waste has not been applicable due to a high $CaCO_3$ content in the shell waste.[38] To address this disposal issue, the ChiBio project was initiated. As a multinational European research consortium the ChiBio project aims at the development of an integrated biorefinery platform to utilize a variety of crustacean shell wastes from different origins such as the EU, Africa and Asia.[38] These shell wastes are converted by the ChiBio biorefinery into chemical intermediates (e.g. functionalized fatty acids and pyrrole derivatives), which resemble precursors for the production of value added bio-based polymers.[38] Biorefining in the ChiBio refinery starts with the application of pretreatment methods to remove $CaCO_3$ and protein residues form the crustacean shell waste.[38,39] The pretreated shells consist of chitin, which is the second most abundant biopolymer after cellulose (Figure 2) and consists of N-acetylglucosamine monomeric units. [40]

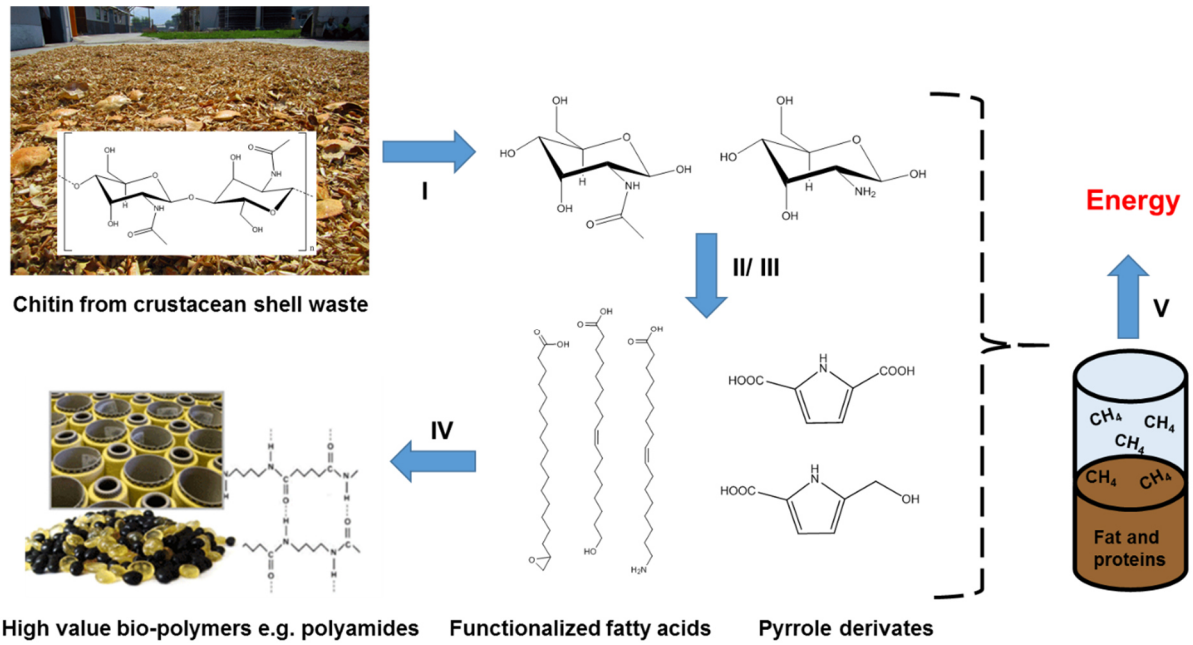


Figure 2.[39] Illustration of the chitin based bio-refinery for the production of speciality chemicals such as bio-polymers. (**I**) Crustacean shell waste is processed into the monomeric sugars N-acetylglucosamine and glucosamine. (**II**) N-acetylglucosamine is converted into functionalised fatty acids by whole-cell biocatalysis, while glucosamine is processed into bi-functionalised pyrroles using an enzymatic route (**III**). Building blocks created from II and III are used for the production of bio-polymers (**IV**). Biomass waste streams produced in the bio-refinery are used for the production of bioenergy from biogas.

In an enzymatic facilitated downstream process, the shells are depolymerized by chitinases into the monomeric sugars N-acetylglucosamine and acetylglucosamine.[38,39] Using enzymatic and whole-cell biocatalytical routes developed by the ChiBio consortium, these monomeric sugars are processed into building blocks for the production of bio-polymers such as polyamides (Figure 2).[39]

Whole cell biocatalytic conversion of N-acetylglucosamine is facilitated by oleaginous yeasts, which utilize N-acetylglucosamine as a carbon source and convert it into cellular lipids.[38] Genetical engineered oleaginous yeast strains convert the shell hydrolysate into lipids comprising non-native and functionalized fatty acids.[38] Suggested examples of chemical modifications on fatty acids are terminal hydroxyl-, carbonyl-, epoxy- or amino groups.[38,39] In a variety of downstream processes, the lipids are separated from the biomass and enzymatically cleaved into glycerol and functionalized fatty acids.[38] Subsequently, the functionalized fatty acids are used as monomer building blocks for bio-based polymers (e.g. polyamides).[38,39] The enzymatic route focused on the development of a process, which converts glucosamine into the pyrrole derivatives 1H-pyrrole-2,5-dicarboxylic acid and 5-hydroxymethyl-1H-pyrrole-2-carboxylic acid for the polymer industry.[38]

Biorefining of crustacean shell waste is accompanied by utilizing bio-waste streams (e.g. residue yeast biomass, glycerol) for the production of bioenergy to improve process efficiency.[38,39] In a life cycle analysis, the environmental impact of the process chain from cradle-to- product is evaluated.[38,39] The ChiBio project is targeting to demonstrate economic viability of the bio-refinery process.

2.1.2 Single Cell Oils

In most microorganism the production of triacylglycerol (TGA) does not exceed 20% of their dry biomass.[41,1,42] However, few examples are known to greatly exceed this threshold and are therefore referred to as oleaginous species. The lipids produced by these species are known as single cell oils (SCOs).[41,43] Oleaginous microorganisms can be found in yeast, molds, algae and bacteria.[1,42,44,45] Some of these species can accumulate lipids up to 80% of their cellular dry weight (DCW).[1] These lipids are enclosed in lipid droplet cell organelles, which can either serve as energy storage to prevent starvation or be used as commodities.[42,43,46] A selection of oleaginous species is illustrated in table 1.

Table 1.[42,44] Lipid content and fatty acid distribution of selected oleaginous microorganisms.

Organism	Lipids (DCW, (*w*/*w*) %)	Major Fatty Acid Residues (Relative % *w*/*w*)											
		14:0	16:0	16:1 n-7	18:0	18:1 n-9	18:2 n-6	18:3 n-3	18:3 n-6	20:4 n-6	20:5 n-3	22:6 n-3	Others
Oleaginous yeasts													
Candida sp. 107	42	Trace	44	5	8	31	9	1	-	-	-	-	-
Cryptococcus albidus	65	Trace	12	1	3	73	12	-	-	-	-	-	-
Lipomyces starkeyi	63	Trace	34	6	5	51	3	-	-	-	-	-	-
Rhodosporodium toruloides	66	18	3	3	66	-	-	-	-	-	-	-	-
Rhodotorula glutinis	72	Trace	37	1	3	47	8	-	-	-	-	-	-
Trichosporon beiglii	45	Trace	12	-	22	55	12	-	-	-	-	-	-
Trichosporon oleaginosus	58	Trace	32	-	15	44	8	-	-	-	-	-	-
Yarrowia lipolytica	36	Trace	11	6	1	28	51	-	-	-	-	-	-
Filamentous Fungi													
Entomophthora coronata	43	31	9	-	2	14	2	-	1	-	-	-	12:0 (40%)
Mortierella isabellina	86	1	29	-	3	55	3	-	3	-	-	-	-
Aspergillus terreus	57	2	23	-	Trace	14	40	21	-	-	-	-	-
Claviceps purpurea	60	Trace	23	-	2	19	8	-	-	-	-	-	12-HO-18:1 (42%)
Microalgae													
Spirulinna platensis	22	8	63	2	-	4	9	-	12	-	-	-	-

Chlorella vulgaris	52	-	16	2	-	58	9	14	-	-	-	-	-
Crypthecodium cohnii	50	16	16	1	-	21	1	-	-	-	-	40	-
Isochrysis galbana	23	12	10	11	-	3	2	-	-	<1	25	11	18:4 (11%)
Thraustochytrids													
Nannochloropsis oculata	45	4	15	22	-	3	1	-	-	4	38	-	14:1 (13%)
Schizochytrium sp.	40	17	32	8	-	5	-	-	-	-	1	28	22:5 (8%)
Bacteria [44]													
Rhodococcus opacus PD630 (gluconate)	76	4.3	25.7	9.5	3.5	22	-	-	-	-	-	-	C15:0 (6.3%),C17:0 (15.4)
Rhodococcus opacus PD630 (octadecane)	39	4.7	41.7	1.4	14.3	37.9	-	-	-	-	-	-	-

The production of SCOs in large scale fermentations is of great industrial interest.[41,47] Although when cultivated on sugars, the manufacturing costs are considered higher compared to traditional plant and animal derived lipids, SCOs can present a financial incentive when used in a biorefinery to produce high-value fats by using waste and by-products from the agro-industrial sector.[41,47] In comparison to plant oils, SCOs have the advantage of being produced in shorter production cycles of several days compared to the seasonal growth cycles.[1,45] Additionally, SCOs result in much higher oil yields as for the production of plant oil solely the crops are utilized leaving most of the plant biomass unused.

Lipids produced from glycerol or sugar based substrates are referred to as *de novo* lipid production, while lipid storage from fatty acids or n-alkanes is described as *ex novo* lipid accumulation.[41] *De novo* lipid accumulation in oleaginous microorganisms starts when an excess in carbon source is accompanied by simultaneous nutrition limitation.[41] Nitrogen limitation is a universal trigger for *de novo* lipid accumulation in yeast, fungi, bacteria and algae.[43,44,45,48] Furthermore, depletion in phosphate, sulfur and iron has been demonstrated in some species to result in lipid production.[41,49,50,51,52] However, compared to nitrogen limitation, these nutritional limitations are often less effective and strain dependent.

De novo lipid production is activated after the exponential growth phase when the cells have exhausted a nutrition source such as nitrogen.[41] Since the entire nitrogen source has been fixed in biomass, the cell growth is reduced and eventually stopped.[41] The carbon surplus is transformed into lipids, leading to a gradual increase of the cellular lipid content and a decrease in carbon source.[42] Once the carbon source has been consumed, the lipid content decreases over time as it is being used as an internal storage to prevent starvation.[42] In contrast, *ex novo*

lipid production is independent of nutritional limitation.[41] It occurs during the growth phase and has been mostly studied by the oleaginous yeast *Yarrowia lipolytica*[41]

Microalgae and Thraustochytrids

Microalgae can be cultivated in barren land when photo reactors are used.[53] Fermentation requires less water and produces significantly more amounts of biomass compared to terrestrial plants.[54] Photoautotrophs microalgae such as *Chlorella vulgaris*, *Spirulinna platensis, Isochrysis galbana*, *Nannochloris sp.* and *Dunalilla tertiolecta* do not require hydrocarbon feeding sources but are dependent on light and CO_2.[42,55] By contrast, thraustochytrids such as *Nannochloropsis oculata* and *Schizochytrium sp.* belong to heterkont algae. They are grown heterotrophically in the presence of salt (NaCl).[42,55,56] Microalgae strains are able to accumulate large quantities of lipids under environmental stress conditions such as nitrogen starvation.[57,58] In photoautotrophic microalgae the absence of nitrogen significantly reduces the protein and chlorophyll synthesis.[59] At the same time the production of triacylglycerol serves as an electron sink and storage for carbon and energy.[60] Members of the microalgae and thraustochytrids family show a high degree of long polyunsaturated fatty acid.[42,61,62] The strains *Crypthecodium cohnii* and *Schizochytrium sp.* are used for the commercial production of docosahexenoic acid (22:6(n-3)).[42] This high value fatty acid is crucial for the neural development and is supplemented to infant formulas.[42,63] Currently, the algae based production of docosahexenoic acid is challenged by the availability of cheap fish oil resources.[42] Under optimal conditions microalgae produce about 4 g/L biomass and 2.8 g/L lipids in a photo bioreactor.[54] The large scale production as substrate for a bio-refinery is currently economically not attractive as costs exceed those for the production of vegetable oil.[64]

Oleaginous Yeasts

Oleaginous yeasts can accumulate up to 80% lipids as dry weight.[1] About 30 strains of over 600 yeast species have been characterized as SCO producers including typical genus *Yarrowia*, *Candida*, *Rhodoturula*, *Rhodosporidium*, *Cryptococcus*, *Lipoymces* and *Trichosporon*.[1] Oleaginous yeasts can generate lipids from various carbon sources and are more flexible utilizing substrates compared to oleaginous fungi, microalgae and bacteria.[45,65] Consequently, glucose, xylose and arabinose from hemicellulose, solids from wheat bran fermentation, dairy serum, molasses, mannitol, glycerol and wastewaters were used for lipid production.[1] The amount of oil produced is strongly influenced by the nitrogen content with respect to the substrate used.[65] In this regard, the lipid contents produced by fermentation of molasses are

comparably low due to high nitrogen contents.[1] When sugars from hemicellulose are utilized as substrate, the hydrolysis is accompanied by the generation of toxic compounds such as furfural, 5-hydroxymethylfurfural (HMF), hydroxybenzaldehyd, vanillin, acetic acid and formic acid that inhibit cell growth.[66,67,68] Consequently, the hydrolysate has to be detoxified before being applied as growth medium. Alternatively, tolerant yeast strains could be used.[66]

Unlike microbial oil from fungi and algae, yeast oil has a simple fatty acid composition containing the chain lengths C16 or C18.[41] The most frequently found fatty acids are palmitic acid (16:0), palmitoleic acid (C16:1), stearic acid (18:0), oleic acid (18:1) and linoleic acid (C18:2). Palmitic acid (16:0) and oleic acid (C18:1) are the most abundant fatty acids found in oleaginous yeasts.[41] However, the fatty acid composition is subject to fluctuations and is dependent on the carbon source, pH, temperature and cultivation method applied.[65] Oleaginous yeasts can be grown to high cell densities. For example *Trichosporon oleaginosus* was grown on glycerol to a biomass of 118 g/L with a lipid accumulation of 25%.[69] In another study, *Rhodosporodium toruloides* was grown on glucose to 106 g/L and produced a total lipid yield of 72 g/L.[70]

Due to the simple fatty acid profile and cheap availability of plant oil, single cell oil from oleaginous yeasts has not been explored commercially. However, when oleaginous yeast strains become available that produce high value fatty acids from low cost substrates, a commercial application is economically viable.[71]

Oleaginous Filamentous Fungi

Filamentous fungi show a greater diversity in the fatty acid composition but display reduced substrate accessibility compared to oleaginous yeasts.[45,65] *Mortierella isabellina* was grown on glucose to a cell density of 36 g/L with a total lipid yield of 18 g/L.[72] The lipid content of filamentous fungi comprises a higher proportion of poly unsaturated fatty acids (PUFs).[42] Consequently, molds have been explored for the commercial production of high value fatty acids such as γ-linolenic acid (18:3 (n−6)) and arachidonic acid (20:4(n-6)).[42] γ-Linolenic acid has been claimed to have beneficial effects on human health including atopic eczema, premenstrual tension and cancer.[42] A plant derived source for γ-linolenic acid is evening primrose oil.[42] In order to obtain a more reliable and cheaper source of γ-linolenic acid, a commercial production was done by using single cell oil from the fungi *Mucor circinelloides* and *Mortierella isabellina*.[42] However, the production was ceased after 6 years due to profit erosion in the production of SCO.[42] Arachidonic acid (20:4(n-6)) is a commercially produced fatty acid by *Mortierella alpine*.[42] Arachidonic acid and docosahexaenoic acid are the

predominate fatty acids in neural tissue and are essential for neural development.[42] Consequently, arachidonic acid and docosahexaenoic acid are introduced into infant formula and produced as a high price oil supplement.[42]

The lipid production in filamentous fungi is mainly influenced by the nature of the nitrogen source, carbon source, temperature, agitation and pH in broth.[65] In analogy to oleaginous yeasts various parameters can affect fatty acid composition.[65]

Oleaginous Bacteria

Bacteria usually accumulate polyhydroxyalkanoates (PHAs) as energy storage.[44,65] Only a few strains produce triacylglycerol, which belong to the actinomyctes group such as *Streptomyces*, *Nocardia*, *Rhodococcus*, *Mycobacterium*, *Dietzia* or *Gordonia*.[44,65] The fatty acid composition of bacteria is quite different from microbial oil. Dependent on the carbon source used for cultivation, bacteria produce unusual fatty acids.[44] For example an *R. opacus* strain PD630 cultivated on phenyldecane incorporates phenyldecaneacid into the lipids, while the same strain cultivated on gluconate produces lipids with uneven carbon numbers.[44] *Streptomyces* species accumulate up to 60% cellular lipids, but produce only small amounts of biomass.[44] When *Rhodococcus opacus* PD630 is grown on gluconate or olive oil, the cells accumulate lipids of 76 and 87%, respectively.[44] Although *Rhodococcus opacus* was grown to high cell densities of 37.5 g/L in a bioreactor[73], lipid production in oleaginous bacteria only play a minor role in single cell oil.

2.1.3 Biochemistry of Lipid Production in Oleaginous Yeasts

The biosynthesis of lipid in oleaginous yeasts is very comparable with low oil producing yeasts such as the baking yeast *Saccharomyces cervisiae*.[1] However, when non-oleaginous species are cultivated under lipid producing conditions, they usually accumulate less than 10% lipids of their dry biomass. By contrast, oleaginous yeasts can store up to 70% as cellular dry weight.[43] It has been demonstrated that oleaginous microorganisms are not in possession of a hyper producing fatty acid biosynthesis, but instead are capable of producing large amounts of acetyl-CoA, the basic unit of fatty acid biosynthesis.[41] Lipid accumulation usually takes place when limitations apply and an excess of carbon source is present.[74] Nutritional limitation is frequently depicted by a lack of a nitrogen, but alternative limitations such the depletion of phosphate, magnesium, zinc, iron and sulfur can trigger lipid accumulation.[74] While nitrogen

limitation yields the highest lipid content and its metabolic characteristics are well known, other limitations are highly strain dependent, poorly characterized and yield a reduced lipid content.[74] When a non-oleaginous yeast encounters a substrate limitation, the remaining carbon source is used for the synthesis of polysaccharides such as starch, β-glucan and mannan, while the lipid biosynthesis is reduced.[1] Under the same conditions an oleaginous yeast channels the excessive carbon source into the lipid biosynthesis and storage.[1,41]

In the case of the well characterized nitrogen limitation, oleaginous yeasts degrade intracellular adenosine monophosphate (AMP) by the AMP-desaminase to yield inosine monophosphate and ammonium.[41,43] Diminishing cytosolic AMP concentration alters the citrate cycle in the mitochondria.[41] This unique effect is only found in oleaginous organisms and is caused by the allosterical activation of the isocitrate dehydrogenase (ICDH) by AMP.[41,43,75] Consequently, the isocitrate dehydrogenase loses its activity, which leads to an accumulation of mitochondrial isocitrate that is rapidly converted into citrate by the isocitrate aconitase (AC) (Figure 3).[41,43,75]

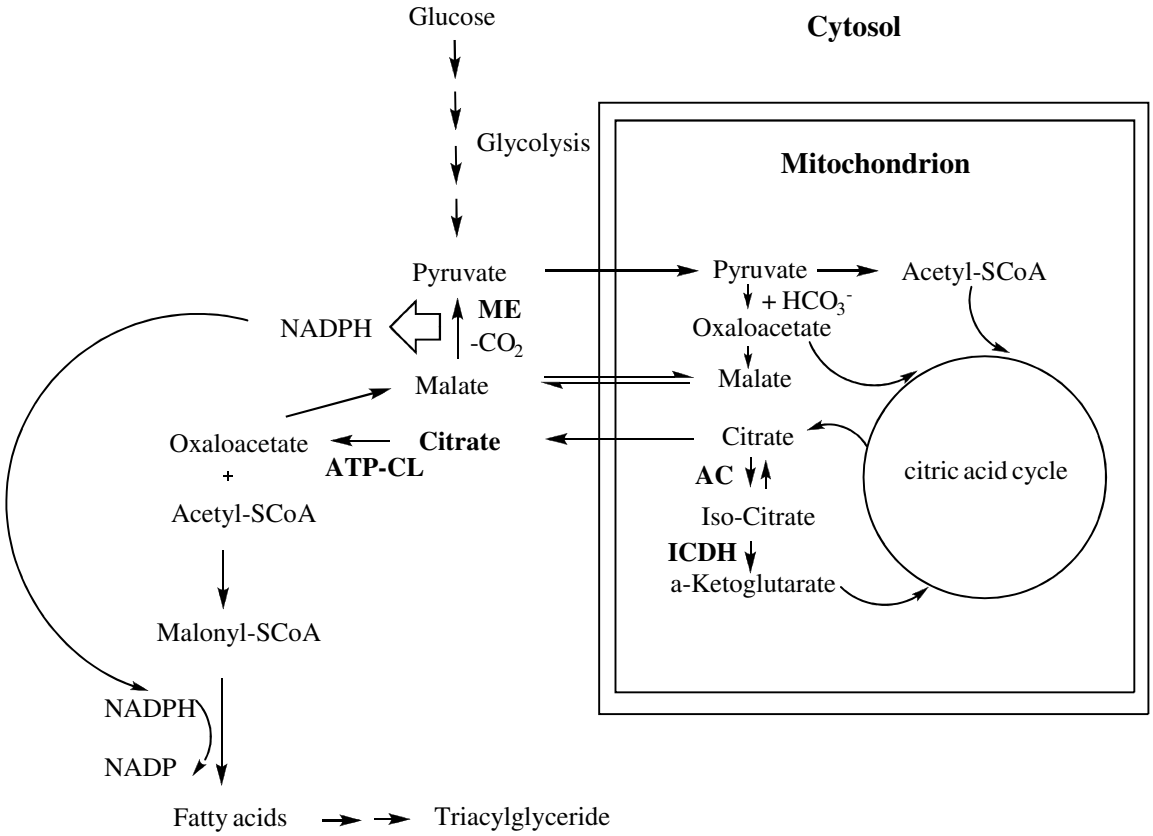


Figure 3. Metabolism in the oleaginous yeast when lipid production is initiated by nitrogen limitation. Enzymes: Ac, acotinase; ACL, ATP-citrate lyase; ICDH, iso-citrate dehydrogenase; ME, malate enzyme. Figure was adapted from [41]

When the mitochondrial citric acid concentration significantly increases, citrate is transported into the cytosol in exchange with L-malate (Citrat-Malat-Shuttle) (Figure 3).[41,43]

In oleaginous yeasts the mitochondrial citrate concentrations are three to four times higher than in non-oleaginous yeasts.[1] Furthermore, the antiporter which causes the citrate flow in the presence of L-malate is more efficient.[1] The cytosolic citric acid is cleaved by the ATP consuming citrate lyase (ACL) into acetyl-CoA and oxaloacetate, which is the second unique step occurring in oleaginous microorganisms (Figure 3).[1,75] No other microorganism has been found that can accumulate more than 20% of its biomass as lipids without ACL activity.[43] It is therefore regarded as one of the key enzymes for the production of SCOs.[43,75]

While acetyl-CoA can immediately be used in lipogenesis to build up palmitic acid, oxaloacetate is further metabolized to L-malate by $NADH/H^+$ consumption.[43,75] Next, L-malate is decarboxylated by malate enzyme (ME) to yield pyruvate, which generates $NADPH/H^+$ an important redox equivalent to the fatty acid biosynthesis.[41,43,75] Alternatively, L-malate may be used by the citrate malate shuttle.[74] The mitochondrial L-malate is then consumed by the citric acid cycle after oxidation to oxaloacetate (Figure 3).[74] Malate enzyme is considered the third essential enzyme to accumulate high amounts of lipids.[41,43,75,76] Fatty acid biosynthesis requires high amounts of $NADPH/H^+$, which in non-oleaginous yeasts is solely provided by the glucose oxidizing phase of the pentose phosphate pathway.[75] When oleaginous microorganisms are cultivated on alternative carbon sources than glucose, malic enzyme becomes the main source of producing $NADPH/H^+$.[76]

For the fatty acid production the cytosolic metabolic bulk chemicals acetyl-CoA and $NADPH/H^+$ are consumed and converted into palmitic acid, which is the primer for longer chain saturated and unsaturated fatty acids.[41,76] The fatty acid biosynthesis called lipogenesis pathway is a cytosolic process catalyzed by the fatty acid synthase - a multi-enzyme protein.[41]

The reaction equation is:[41]

acetyl-CoA + 7 malonyl-CoA + 14 $NADPH/H^+$ -> palmitic acid + 7 CO_2 + 8 CoA + 14 $NADP^+$ + 6 H_2O.

2.1.4 Fatty Acid Desaturases and Elongases

Desaturases

Unlike microbial oil from fungi and algae, yeast oil has a simple fatty acid composition containing the chain lengths C16 or C18.[65] The most frequently found fatty acids are palmitic acid (16:0), palmitoleic acid (C16:1), stearic acid (18:0), oleic acid (18:1) and linoleic acid (C18:2).[41]

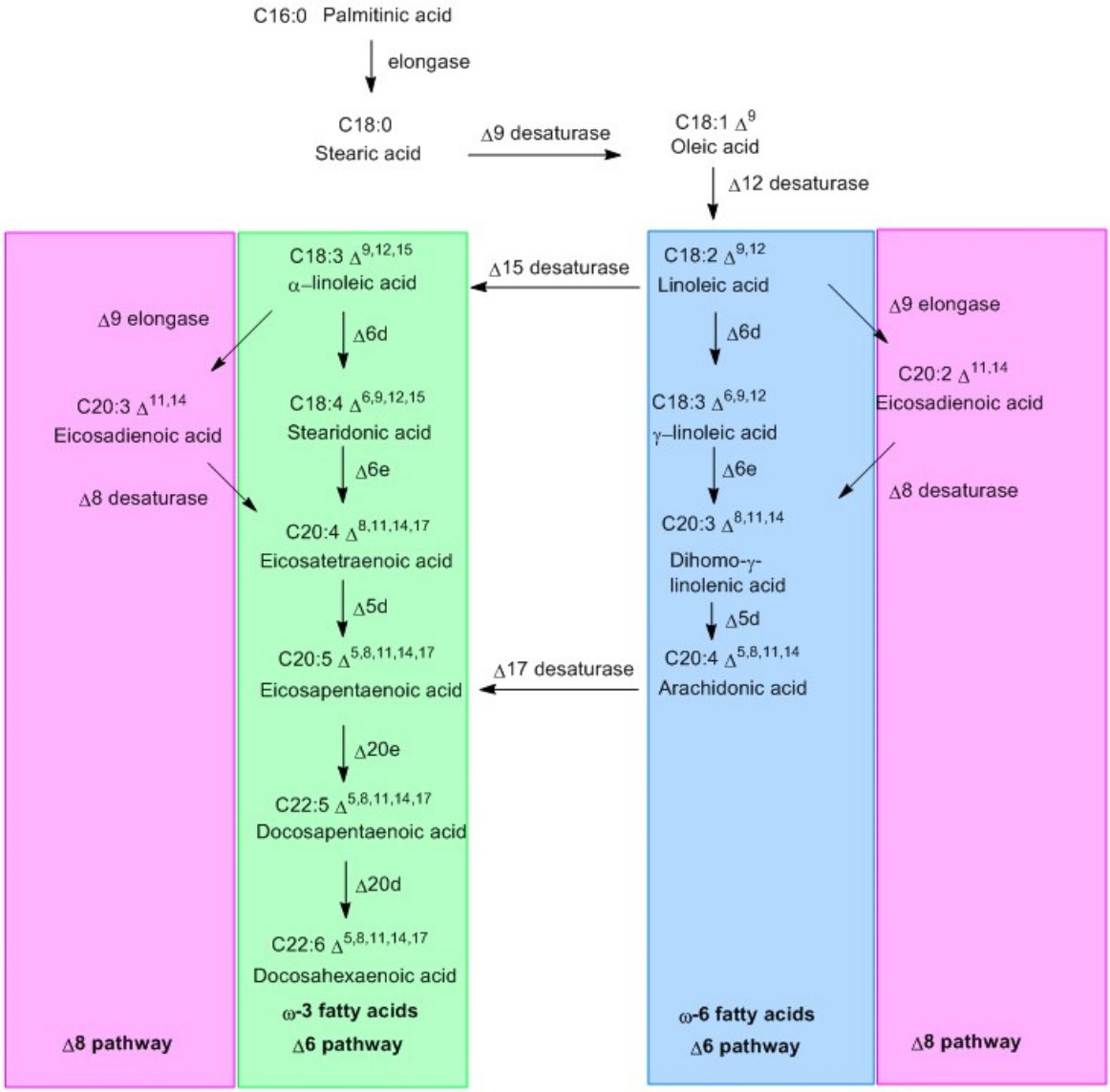


Figure 4. Biosynthesis of ω-3 (green) and ω-6 (blue) very long chain polyunsaturated fatty acids. The Δ6 pathway (green and blue) is characterized by initial Δ6 desaturation and Δ6 elongation steps, whereas the Δ8 pathway (magenta) consists of an initial Δ9 elongation step followed by Δ8 desaturation. Adapted from [43,77]

Fatty acids other than palmitic acid are created by a sequence of desaturases and elongases enzyme activities illustrated in figure 4.[43,77] In contrast to the fatty acid biosynthesis, chain elongation takes place in the endoplasmic reticulum using fatty acid-CoA esters and malonyl-CoA as substrate.[78]

Desaturases catalyze the introduction of double bonds.[78] They belong to the group of metalloenzymes that use non-heme di-iron active sites to activate molecular oxygen and abstract two hydrogens from fatty acids, resulting in the formation of a double bond and the release of water.[78] The oxygen dependency of desaturases can be used for shifting the fatty acid profile to more saturated fatty acids by applying oxygen limitation during the fermentation.[42] Desaturases can be classified according to the type of fatty acid esters that are accepted as substrate.[78] Fatty acids can be present as acetyl-CoA ester, be connected to the sn-2 position of phosphatidylcholine or be attached to the acetyl-carrier protein (ACP).[78] Desaturase that utilize acetyl-CoA esters are present in animals, yeast and fungi.[78] Acetyl-phosphatidylchloine esters are metabolized by desaturases found in plants, fungi, and cyano bacteria. Acyl-ACP desaturases are found in plant plastids as soluble proteins.[78] The Δ9 fatty acid desaturase converts stearic acid into oleic acid and is the only desaturase in yeast that uses the stearic acid-CoA ester as substrate.[78] In contrast, the Δ12 desaturase converts the phosphatidylcholine ester of oleic acid into linoleic acid.[78]

Elongases

Elongases are using malonyl-CoA to extend fatty acid up to a chain length of 24 carbon atoms in a similar way than the fatty acid synthesis.[79] For the elongation process, fatty acids acyl-CoA esters are used as substrate.[80] Fatty acids with a chain length >18C are called very long chain (VLC) fatty acids.[43] The biosynthesis of VLC fatty acids falls into two common categories, the n-3 and n-6 series dependent on the position of the final double bond (Figure 4).[43] This classification is less compelling and several links between both pathways exist. The n-3 and n-6 series are derived from the Δ6 pathway.[43] An alternative Δ8 pathway for biosynthesis of 20-carbon poly unsaturated fatty acids (PUFAs) is used by euglenoids that lack the Δ6-desaturase activity (Figure 4).[78,81] In this pathway the linoleic acid and α-linoleic acid are first elongated to ecosadienoic and ecosatrienoic acid, which are subsequently desaturated by a Δ8 desaturase to dihomo-γ-linolenic acid and eicosatetraenoic acid.[81,82]

Many algae produce very long chain polyunsaturated fatty acids (VLCPUFAs) as major fatty acid components.[63] Since algae are the beginning of the marine food chain, these fatty acids are enriched in other marine species.[63] VLCPUFAs such as eicosapentaenoic acid (EPA; 20:5 $\Delta^{5,8,11,14,17}$), docosahexaenoic acid (DHA; 22:6 $\Delta^{4,7,10,13,16,19}$) and eicosapentaenoic (EPA, C20:5) acid are essential for human health and of great industrial interest.[5] However, they are predominantly derived from marine oils, placing enormous pressure on diminishing marine

resources.[83] The development of alternative sustainable sources for VLCPUFAs that are economically competitive could be a chance to preserve marine eco system.[5]

2.1.5 Conjugated Linoleic Acids

Conjugated linoleic acids (CLAs) describes linoleic acid (LA) isomers with conjugated double bonds such as *cis*-9, trans-11 CLA, *trans*-10, *cis*-12 CLA and *trans*-9, *trans*-11 CLA.[84] Natural sources of CLA are meat and dairy products derived from ruminants, which are obtained as minor fractions of the total lipid content.[85] The primary source of CLAs are bacterial origins and are produced by strains such as *Lactobacillus reuteri*, *Clostridium sporogenes* and *Propionibacterium acnes*.[86,87] These bacteria produce the cytosolic enzyme called conjugated linoleic acid isomerase, which is restricted to utilize free linoleic acid as substrate.[87] Mechanistically, isomerization of linoleic acid is facilitated by a FAD mediated intramolecular hydride transfer without the consumption of cofactors.[88] CLAs display several beneficial pharmacological effects (ex. anti cancer).[87] To satisfy high demand, CLAs are produced chemically by alkaline isomerization of linoleic acid, a process that is energy intensive and lacks selectivity.[87]

2.1.6 Biosynthesis of Triacylglycerol

For the lipid assembly, fatty acid CoA esters are linked to glycerol.[41] In the first step glycerol-3-phosphate (G3P) is acylated by G3P acyltranferase (GAT) at the sn-1 position to yield 1-acyl-G3P (lysophosphatidic acid-LPA).[41] The second acylation step at the sn-2 position is catalyzed by the lysophosphatidic acid acyltransferase to yield phosphatidic acid (PA).[41] Next, PA is dephosphorylated by the phosphatidic acid phosphohydrolase (PAP) releasing diacylglycerol (DAG), which is eventually acetylated by the diacylglycerol acyltransferase to produce triacyl glycerol (TGA).[41]

2.1.7 The Oleaginous Yeast *Trichosporon Oleaginosus*

Trichosporon oleaginosus ATCC 20509 is an oleaginous yeast and has recently been assigned to the basidomycetous genua *Trichosporon.*[2] It has formally been known as *Cryptococcus curvatus*, *Candida curvata* and *Apiotrichum curvatum* and was extensively studied for its ability to produce up to 70% biomass as triglycerides resembling a cocoa-butter-like fatty acid composition.[1,2,3]

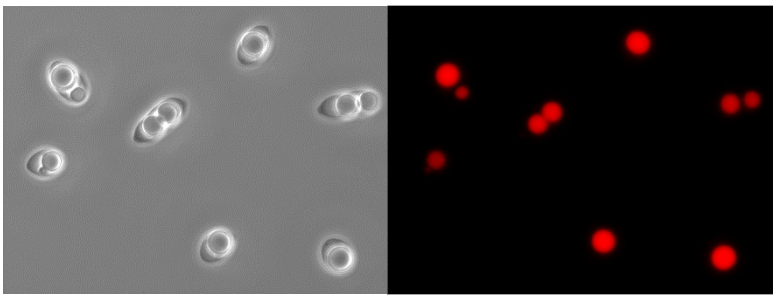


Figure 5. Fluorescence microscopic picture of wild type *T. oleaginosus* cells. Cultivation was done for 7 days in glucose containing medium with nitrogen limitation (48% lipids of their dried cell weight). The cells were stained with Nile Red to visualize the lipid bodies. Transmitted microscopy is shown on the left side, while the fluorescence is illustrated on the right side.

T. oleaginosus can be grown in high cell density cultures from various waste materials such as whey permeate, crude glycerol, sweet sorghum bagasse, pectin-derived carbohydrates, wheat straw hydrolysate and N-acetylglucosamine from chitin hydrolysate[3,67,69,89,90,91,92] Its ability to utilize highly diverse carbon sources for lipid production distinguishes *T. oleaginosus* from other oleaginous yeasts. Additionally, lipid and biomass production are mostly unaffected by fermentation inhibitors such as furfural, which are constituents of waste biomass hydrolysates.[67] Consequently, *T. oleaginosus* is an ideal yeast strain for the production of high value lipids from cost efficient waste biomass hydrolysates or alternative biotechnological waste streams such as crude glycerol from biodiesel production.

So far a series of *T. oleaginosus* mutants have been created by random UV-based mutagenesis or the chemical mutagen N-methyl-N'-nitro-N'-nitrosoguanidine (MNNG) to increase the stearic acid content to resemble a more cocoa-butter like fatty acid profile.[93,94,95] Starting from a $\Delta 9$ desaturase mutant that could only be grown when oleic acid was supplemented, revertant mutants could be created that had only a partially active $\Delta 9$ desaturase.[93] Promising mutants were R25.75 and R26-17 as shown in table 2.[93] R26-17 showed the closest fatty acid composition compared to cocoa-butter.[93] In a second approach partial $\Delta 9$ desaturase activity

was regained by intraspecific spheroplast fusion of a methionine auxotrophic mutant with a Δ 9 desaturase mutant leading to strain F33.10.[95] The latter displayed an almost perfect cocoa-butter like fatty acid distribution (Table 2).

Table 2. *T. oleaginosus* mutants created for the production of cocoa-butter like single cell oil. [93,95]

	Relative % (*w*/*w*) of fatty acids				
	C16:0	**C18:0**	**C18:1**	**C18:2**	**18:3 (n-3)**
Wild Type *T. oleaginosus*	28	14	44	10	1.1
Cocoa butter	30-23	37-32	37-30	4-2	-
T. oleaginosus R25.75	29	30	28	8	1
T. oleaginosus R26-17	22	54	12	6	-
T. oleaginosus R33.10	24	31	30	6	4

Despite these remarkable strain development efforts that demonstrate the engineering potential of *T. oleaginosus* to produce altered fatty acid compositions, a transformation system and heterologous gene expression has not been demonstrated to date.

2.1.8 Scope of This Work

The research described in this chapter is part of the ChiBio project and covers genetic engineering approaches of the oleaginous yeast *T. oleaginosus* to convert biomass derived feedstocks such as N-acetylglucosamine, glucose and xylose into high value lipids.

T. oleaginosus displays a high substrate flexibility and is known to produce considerable lipids from N-acetylglucosamine when phosphate limitation is applied.[92] Furthermore, the yeast strain tolerates higher nitrogen concentrations in the fermentation medium without adverse effects on *de novo* lipid biosynthesis compared to other oleaginous yeast strains. Consequently, *T. oleaginosus* is an ideal fermentative platform organism for the production of modified lipids and was chosen for this study. In the concept of the Chibio bio-refinery, whole cell biocatalysis is used to convert crab shell hydrolysate to functionalized lipids that serve as precursors for bio-polymers.[38]

This study focuses on establishing genetic accessibility of *T. oleaginosus*. Furthermore, the ability of *T. oleaginosus* to perform as a flexible production platform for the recombinant generation of tailor made lipids derived from various monomeric sugars that are commonly accessible by biomass waste streams, is evaluated.

Therefore, the heterlologous gene expression of the bi-functional desaturase Fm1[5,77] from *Fusarium moniliforme* and the elongase IgASE2[82] from *Isochrysis galbana* is evaluated in *T. oleaginosus* for the production of the very long chain fatty acids eicosadienoic acid (EDA, C20:2 $\Delta^{11,14}$) and eicosatrienoic acid (ETE, C20:3 $\Delta^{11,14,17}$). These fatty acids are important metabolic intermediates towards the bio-synthesis of eicosapentaenoic acid (EPA; 20:5 Δ5,8,11,14,17) and docosahexaenoic acid (DHA; 22:6 Δ4,7,10,13,16,19) from marine oil resources.[4,5] Furthermore, the enzymatic accessibility of free intracellular fatty acids in *T. oleaginosus* is assessed by the heterologous expression of the linoleic acid isomerase PAI[87] from the bacteria *Propionibacterium acnes*. This enzyme converts linoleic acid into *E*-10, *Z*-12 conjugated linoleic acid (CLA), which is a pharmaceutically important fatty acid.[6,87]

Last but not least, the recombinant generation of tailor made lipid by the genetically engineered *T. oleaginosus* strains is evaluated on various monomeric sugars such as acetylglucosamine, glucose and xylose.

2.2 Results and Discussion

2.2.1 Development of a Transformation Procedure for *T. oleaginosus*

Selection Marker

At the beginning of this study a genetic transformation system for *T. oleaginosus* was not available. To develop a transformation tool, a suitable selection system to distinguish between transgenic and wild type cells was required. A commonly used selection system for yeast and filamentous fungi is based on the dominant selection marker hygromycin B from the bacteria *Streptomyces hygroscopicus* that prevents cell growth by stabilizing the tRNA-ribosomal acceptor site.[96,97,98,99,100,101,102,103,104,105]. To circumvent the lethal effect of hygromycin B, a successfully transformed yeast cell must express the resistance gene *hph* that encodes the hygromycin B phosphotransferase and inactivates hygromycin B by phosphorylation.[103] It was found that *T. oleaginosus* is sensitive to hygromycin B at concentrations of 100 µg/mL (YPD medium). Consequently, the transformation system was based on hygromycin B as dominant selection marker.

Promotor and Terminator Used for the Transformation System

To enable strong expression of hygromycin B resistance gene, the constitutive promotor of the glyceraldehyde-3-phosphate dehydrogenase (GPD, EC1.2.1.12) gene was chosen. As part of the glycolysis the promotor is considered as strong, constitutive and its mRNA transcripts can reach 2–5% of the total poly(A)+RNA pool in *Saccharomyces cerevisiae*.[102] Furthermore, the GPD promotor has been used in the transformation of numerous fungal species and has been exploited for the heterologous gene expression.[102,106,107,108,109] By the time the transformation system was developed, only a 390 bp fragment of the *T. oleaginosus* GPD promotor was available as a NCBI genebank entry. It was questionable whether the truncated promotor sufficiently drives gene expression of the resistance gene. Nevertheless, this was put to a test and a gene expression cassette was created consisting of the 390 bp GPD promotor segment, the hygromycin B resistance gene *hph* and the tryptophan synthase (TrpC) terminator derived from the filamentous fungi *Aspergillus nidulans*.

2.2.2 Transformation of *T. oleaginosus* by the LiAc/SS Carrier DNA/PEG Method and Electroporation.

To date no self-replicating plasmids are known for *T. oleaginosus*. Consequently, the gene cassette was flanked with homologous gene segments (each about 1000 bp) from the ribosomal DNA to facilitate genomic integration by homologous recombination. The rDNA is particularly suitable as an integration target as the rDNA consists of multiple genomic copies, which are easily accessible due to frequent transcription.[110] As a result the rDNA based genomic transformation has been frequently used to genetically engineer yeasts.[111,112,113] Cloning was done using the pRF-HU2 plasmid, which was originally developed for the gene deletion by homologous recombination in the filamentous fungi *Fusarium graminearum*.[101] For transformation different techniques were tested such as transformation by the LiAc/SS carrier DNA/PEG method and electroporation.[114,115] The protocol according to Gietz et al. was followed for the LiAc/SS carrier DNA/PEG method.[114] Optimization was carried out by changing single parameter at a time such as the DNA concentration, heat shock incubation time and the amount of single stranded carrier DNA applied. Parameters varied for the electroporation attempts were either the electrical field or the amount of linear DNA applied. Electroporation was done according to Thompson et al.[115] Unfortunately, neither of these methods and parameters tested was suitable to accomplish a successful transformation.

2.2.3 Transformation of *T. oleaginosus* by *Agrobacterium tumefaciens*

In the next step the *Agrobacterium tumefaciens* mediated transformation was evaluated. *A. tumefaciens* has been reported for the genomic integration of heterologous genes in other non-conventional fungi and yeasts.[102,116,117,118,119] As a plant pathogen *A. tumefaciens* causes tumour crown gall disease by semi-random insertion of small segments of DNA (transfer DNA (T-DNA)) from a tumour inducing plasmid (ti-plasmid) into the plant genome, [119] The T-DNA encodes genes for the opine synthesis, which are synthesised by the plant tumour cells and serve as a nutrition source for the *Agrobacterium*.[120] For the development of the *T. oleaginosus* expression cassette the plasmid pRF-HU2 was used as backbone. Since this plasmid was originally developed for the *Agrobacterium* mediated transformation of filamentous fungi, the backbone of the plasmid could be used without further modification.[101] The pRF-HU2 plasmid (Figure 6) contains a broad host replication initiation gene (TrfA) and the corresponding

replication of origin (oriV), which facilitates self-replication of the plasmid in *E.coli* and A. *tumefaciens*.[101] Furthermore, the plasmid contains a left boarder (LB) and right boarder (RB) oligonucleotide sequence, which confine the T-DNA segment that is recognized by the *Agrobacterium*.[101]

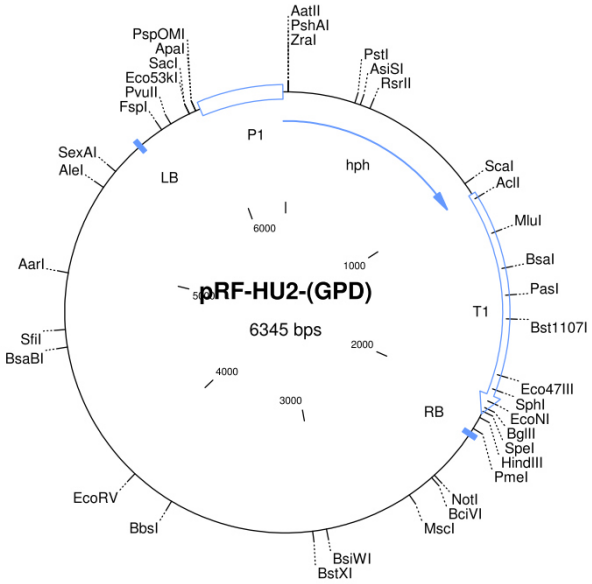


Figure 6. Plasmid map of the pRF-HU2-(GPD) vector displaying P1 (*T. oleaginosus* GPD promotor 390 bp), T1 (tryptophan (TrypC) terminator from *Aspergillus nidulans*), *hph* (hygromycein B resistance gene), LB (left border) and RB (right boarder).

It was tested if the rDNA flanked expression cassette could be integrated into the rDNA of *T. oleaginosus* by *A. tumefaciens* mediated transformation. In addition to the flanked gene cassette a non-directional integrative approach was evaluated. Therefore, the flanked rDNA sequences were removed, which facilitates a random genomic integration by non-homologous recombination. *Agrobacterium tumefaciens* mediated transformation is a far more complicated procedure compared to classical transformation techniques.[119] The procedure consists of two phases.[101,102,119] In the first phase yeast and *A. tumefaciens* containing the modified pRF-HU2 plasmid are mixed and co-cultivated on a Hybond-N^+ membrane, which is placed on top of so called "induction medium agar plates" (IMAS plates).[101,102,119] IMAS plates contain a low pH (between 5-6), a shortage of nutrition and are supplemented with the plant pheromone acetosyringon to trigger the virulence of the *Agrobacterium*.[101,102,119] After several days the second phase starts.[101,102,119] The membrane is transferred to selection agar plates, which consist of typical yeast media (YPD-Media) supplemented with the antibiotic cefotaxime to kill the *Agrobacterium* cells as well as hygromycin B to select for transformed yeast cells.[101,102,119] The second phase takes up to one week.[101,102,119] A lot of parameters including the pH, concentration of acetosyringon, temperature, glucose and glycerol content as well as

incubation times were optimized. Furthermore, the *A. tumefaciens* strains LBA4404 and the hyper-virulent strain AGL1 were tested. The key to a successful *Agrobacterium* mediated transformation of *T. oleaginosus* was the application of the AGL1 strain and using 0.5% (*w/v*) glycerol as sole carbon source in the IMAS plates (Figure 7).

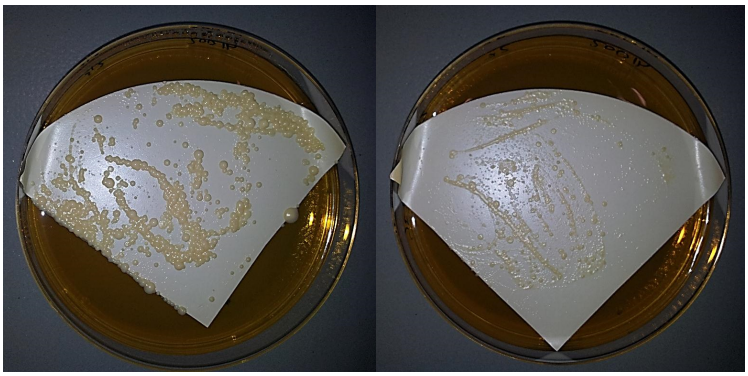


Figure 7. Selection agar plates of an *Agrobacterium tumefaciens* mediated transformation of *T. oleaginosus*. Successful transformants on the Hybond-N+ membrane are shown on the left side. A no template negative control is shown on right side.

Adding additional glycerol or glucose to the medium ultimately resulted in an overgrowth of *T. oleaginosus* cells, which prevented transformation. Furthermore, the best transformation efficiency was obtained with co-cultivation for 48 hours at 24°C using IMAS plates supplemented with 200 µg/mL acetosyringon at pH 5.6. Interestingly, transformation was only established with the plain gene cassette in absence of flanking rDNA sequences for the homologous recombination. This finding is surprising, since it was expected that the rDNA targeted transformation displays equivalent or slightly increased transformation efficiency due to the availability of homologous recombination as alternative pathway for genomic integration. The reason for this may be caused by the formation of unfavourable secondary structure of the T-DNA and its premature degradation.

The *hph* gene could be amplified from isolated genomes of potentially transformed *T. oleaginosus* cells. Furthermore, the transgenic yeast strains were tested for their resistance against different hygromycin B concentration. Interestingly, a wide distribution was observed. Most of the transformed *T. oleaginosus* clones were able to sustain up to 100 µg/mL hygromycin B, while a fraction of clones were able to sustain up to 200 µg/mL hygromycin B. Since ATMT causes random integration in the genome[119], differences in the hygromycin B resistance are most likely caused by variable genomic copy numbers and loci.

2.2.4 Heterologous Expression of a Reporter Gene in *T. oleaginosus*

For expression of the hygromycin B resistance encoding gene (hph) a truncated version (390 bp) of the GPD promotor was sufficient. However, the transgenic *T. oleaginosus* clones only showed a low resistance against hygromycin B, which might be caused by using a truncated promotor segment as well as a non-codon optimized variant of the hph gene. For a high level expression codon optimized genes as well as a strong promotor are essential. While the research for *T. oleaginosus* was conducted detailed genome and transcriptome data for a *T. oleaginosus* strain was kindly provided by the research group of Prof. Dr. Robert Kourist.[121] Using this data enabled us to obtain comprehensive information on the glycerol aldehyde dehydrogenase gene, which comprised the full length GPD promotor and terminator. Consequently, a codon optimized version of the YFP reporter protein was created based on codon usage of the glycerol aldehyde dehydrogenase gene. The codon usage of *T. oleaginosus* was briefly described in literature from cloning of the Δ9 desaturase.[122] In accordance with this data[122] the codon usage contained a high GC content of 62% and was closely related to the fungus *Schizophyllum commune*. For the yfp gene cassette, a 800 bp version of the GPD promotor and a 600 bp segment of the GPD terminator was tested. The yfp gene cassette was introduced in the pRF-HU2(GPD) plasmid and yielded the plasmid pRF-HU2-(GPD)-YFP (Figure 8). After transformation of *T. oleaginosus* with *A. tumnefaciens* harboring the pRF-HU2-(GPD)-YFP plasmid, transformants were screened by microtiter plate reader. Screening analysis revealed a wide distribution of the YFP fluorescence strength. Consequently, a screening approach was used in this study for all transgenic yeast strains. Genomic integration was mitotically stable and the YFP fluorescence levels remained constant after multiple sub-cultivations in absence of the selection marker. The growth kinetics of the wild type strain were compared to the yfp expressing strain and showed no limitations (Appendix Figure 17). The fluorescence microscopy images of *T. oleaginosus* expressing yfp are shown in figure 8. For this figure, the strain was cultivated for 48 hours in rich medium (YPD) and additionally stained with Nile Red to visualize the lipid bodies. Lipid quantification indicated that *T.* oleaginosus accumulates about 10% lipids per dry cell weight (DCW) in nitrogen rich media without nutritional limitation. It is clearly visible that YFP is comprised in the cytosol and covers the entire yeast cell.

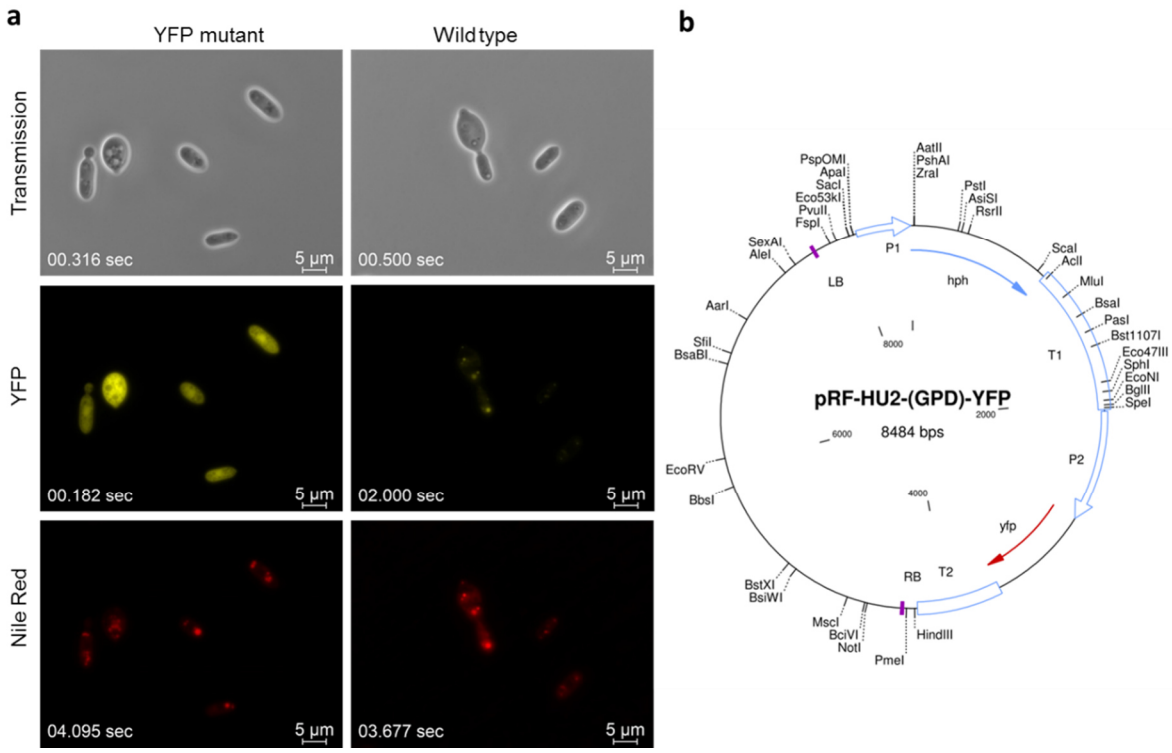


Figure 8. (a) Fluorescence microscopy of YFP overexpressing *T. oleaginosus* and wild type strain. Cells were cultured in rich medium (YPD) for 48 hours and stained with Nile Red to visualize the lipid bodies. (**b**) Plasmid map of pRF-HU2-(GPD)-YFP displaying P1 (*T. oleaginosus* GPD Promotor 390 bp), T1 (tryptophan terminator from *A. nidulans*), *hph* (hygromycein B resistance gene), LB (left border), RB (right boarder), P2 (*T. oleaginosus* GPD promotor 800 bp), T2 (*T. oleaginosus* GPD terminator 600 bp) and *yfp* (yellow fluorescence protein gene).

In general the YFP fluorescence was unexpectedly high in the average of screened colonies. Surprisingly, the best YFP producing strains were detectable to the eye in comparison to the wild type when cultured on agar plates. Consequently, the size of the GPD promotor and terminator was left unchanged for the expression of alternative genes.

2.2.5 Engineering *T. oleaginosus* for the Production of VLC-PUFAs

After the successful production of the YFP protein demonstrated the feasibility to engineer *T. oleaginosus* for the heterologous gene expression, genes to produce non-native fatty acids were evaluated. Therefore, the pathway for the production of the very long chain fatty acids eicosadienoic acid (EDA, C20:3 $\Delta^{11,14}$) and eicosatrienoic acid (ETE, C20:3 $\Delta^{11,14,17}$) was reconstituted, which are important intermediates towards the synthesis of EPA and DHA.[4,5] The total fatty acid content (TFA) of *T. oleaginosus* is dominated by oleic acid, which constitutes about 40-50% and is largely unaffected by cultivation conditions. In contrast, the linoleic acid content is significantly dependent on the growth medium used for cultivation. In the nitrogen rich YPD medium, linoleic acid reaches up to 45%. (TFA) and declined to 5 -10% (TFA) when minimal medium was used with nitrogen or phosphate limitation.

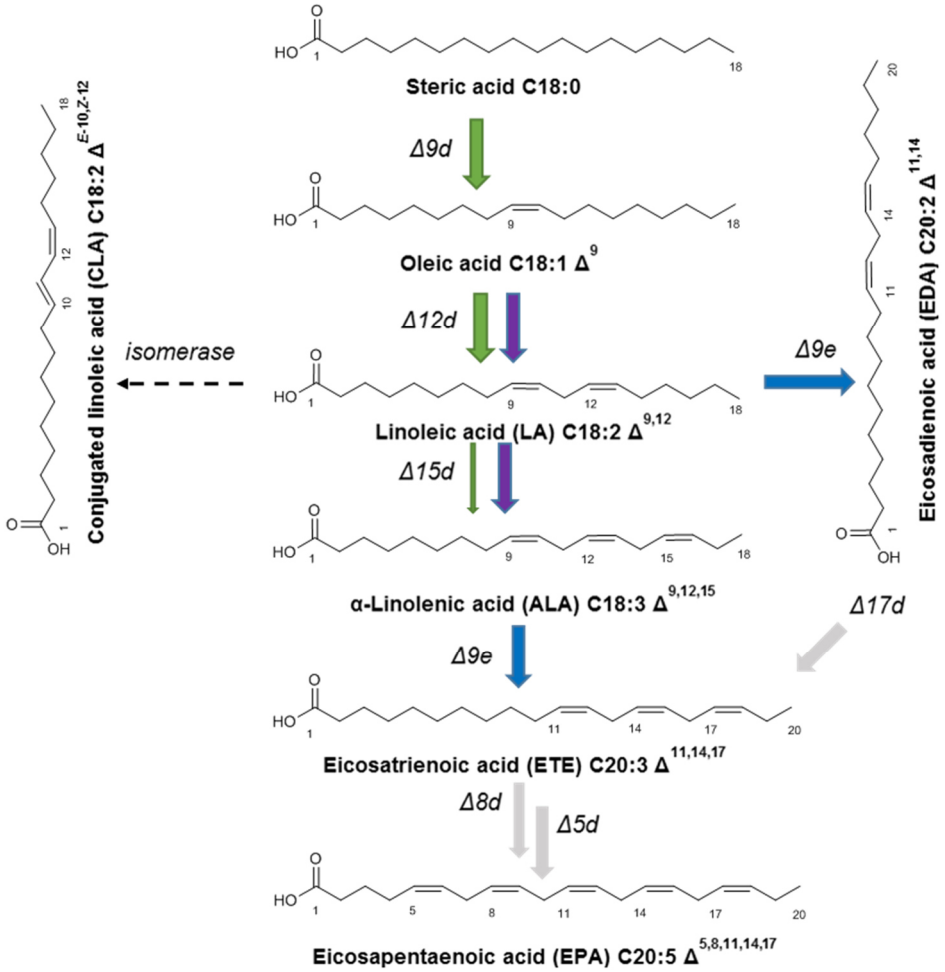


Figure 9. Biosynthetic pathways for the production of eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA) (**grey**). The native pathway in *T. oleaginosus* is indicated in **green**. The bi-functional desaturase Fm1 from *Fusarium moniliforme* (**purple**) converts oleic acid into α-Linolenic acid (ALA), which is then converted into eicosadienoic acid (EDA) and eicosatrienoic acid (ETE) by the elongase IgASE2 from *Isochrysis galbana* (**blue**). Free intracellular linoleic acid is converted into conjugated linoleic acid (CLA) by the isomerase PAI from *Propionibacterium acnes* (**dotted**).

Based on this large consistent oleic acid pool, methodologies to redirect fatty acid biosynthesis towards non-native products were conducted as shown in figure 9. The biochemical pathway of *T. oleaginosus* for the production of fatty acids is illustrated by green arrows (Figure 9). Beside the production of linoleic acid, analysis of the native *T. oleaginosus* fatty acid profile indicated that small amounts of α-linolenic acid (less than 3% TFA) are produced (see Figure 10, below), which has been incorrectly reported as γ-linolenic acid.[92] For the production of eicosatrienoic acid, α-linolenic acid is elongated by a Δ9 elongase (Figure 9 blue pathway).[4,5]

An alternative path for the production of eicosadienoic acid from linoleic acid is the application of the Δ9 elongase (blue pathway) and Δ17 desaturase (grey pathway). Eicosatrienoic acid can be converted to eicosapentaenoic acid by two subsequent Δ8 and Δ5 desaturation steps (Figure 9, grey pathway) and is therefore an important metabolic intermediate.[4,5]

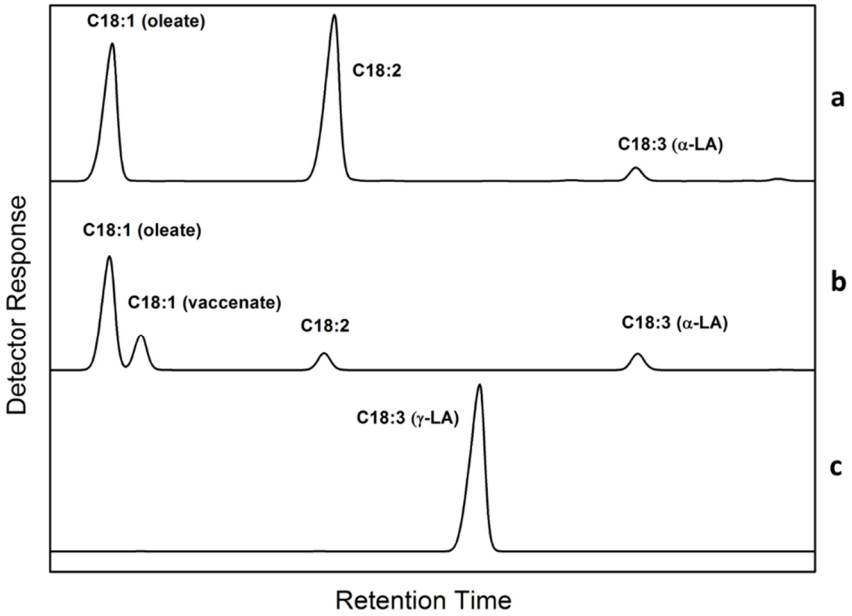


Figure 10. GC-FAME chromatograms of fames derived from wild type *T. oleaginosus* cultivated in YDP medium for 3 days. (**a**), Marine Fames Mix (Restek Corporation) (**b**), γ- Linolenic acid (Sigma Aldrich) (**c**)

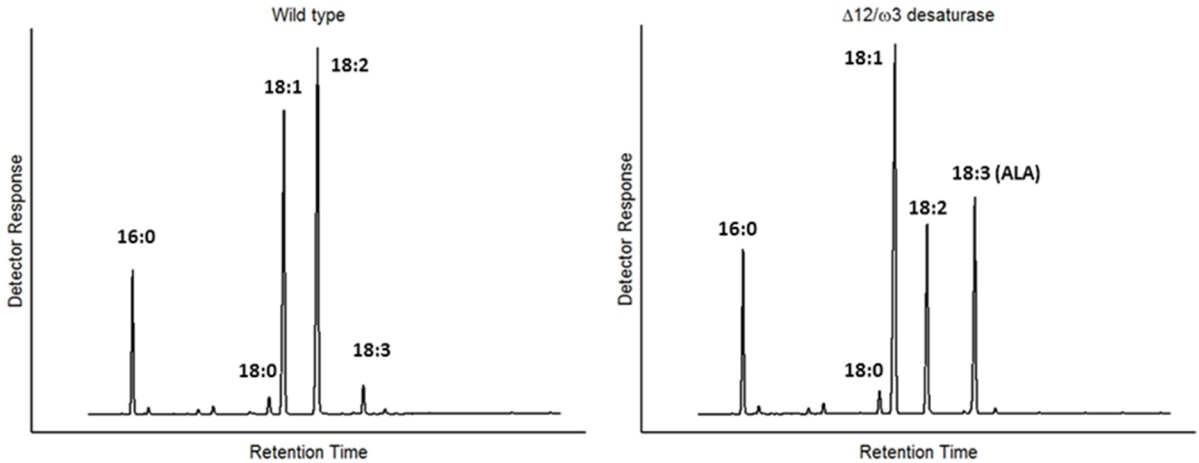
Engineering *T. oleaginosus* for the Production of α-linolenic acid

Since α-linolenic acid is essential for the production of eicosatrienoic acid, the first engineering approach focused on increasing the fatty acid pool of α-linolenic acid. α-linolenic acid is produced by desaturation from linoleic acid at the ω-3 position. Since the precursor linoleic acid content is significantly reduced in lipid production conditions, the metabolic flux from oleic acid to linoleic acid has to be increased as well. The bi-functional Δ12/ω3 fatty acid desaturase (Fm1) of the filamentous fungus *Fusarium moniliforme* was chosen for this task.[5,77] As bi-functional enzyme Fm1 is able to convert oleic acid into linoleic, which is then

subsequently converted in to α-linolenic acid (see figure 9 purple pathway). Fm1 was successfully applied in both *Saccharomyces cerevisiae* and the oleaginous yeast *Y. lipolytica*, which increased the possibility for a successful reconstitution.[5,77]

Membrane protein topology prediction method based on a hidden Markov model (TMHMM) of Fm1 predicted typical five transmembrane helices motives, indicating that the protein is active in the endoplasmic reticulum.[123] A codon optimized version of Fm1 was introduced in *T. oleaginosus*. The fatty acid composition of 30 transformants was analyzed to find the best α-linolenic acid producer.

The fatty acid composition was analyzed after three days growth in YPD in absence of hygromycin B. The transgenic *T. oleaginosus* strain accumulated 21% α-linolenic acid (TFA), which is about one third lower compared to the heterologous expression of Fm1 in the *Y. lipolytica* (28% α-linolenic acid (TFA)) (Figure 11).[77] At the same time, the linoleic acid content decreased in *T. oleaginosus* from 46% to 20% (TFA), compared to a reduction from 39% to 7% (TFA) observed in *Y. lipolytica*.[77]



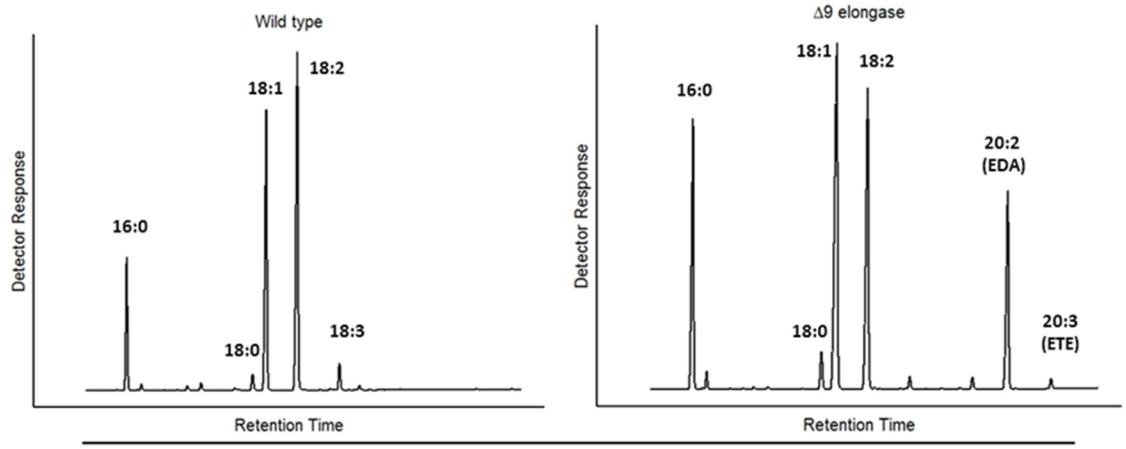
Fatty Acid [% (TFA)]	C16:0	C18:0	C18:1	C18:2	C18:3 (ALA)
Wild type	13.5 ± 0.7	1.7 ± 0.6	35.7 ± 0.2	46.2 ± 1.8	2.8 ± 0.5
Δ12/ω3 desaturase	14.5 ± 1.8	2.3 ± 0.0	42.5 ± 0.4	19.7 ± 1.1	21.1 ± 1.6

Figure 11. GC-FID chromatograms and corresponding fatty acid distribution of a transgenic *T. oleaginosus* strain expressing the Δ12/ω3 desaturase Fm1 from *F. moniliforme*. Cultivation was done in YPD medium for three days.

Engineering *T. oleaginosus* for the Production of eicosatrienoic acid (C20:3) and eicosadienoic acid (C20:2)

In the next step the functional reconstitution of a Δ9 elongase was tested to facilitate the conversion of α-linolenic acid to eicosatrienoic acid as well as from linoleic acid to eicosadienoic acid. A recently discovered Δ9 elongase from the euglenoid *Isochrysis galbana* H29 was selected due to the fact that IgASE2 possessed the highest conversion rate of linoleic acid and α-linolenic acid as substrates, when compared to other Δ9 PUFA elongases.[82] As transmembrane protein, IgASE2 contains five transmembrane α-helices and is integrated in the endoplasmic reticulum.[82] The elongase was reported to be successfully reconstituted in *S. cerevisiae,* providing evidence that a functional reconstitution in yeast is possible.[82] Activity in *S. cerevisiae* was evaluated by supplementing either linoleic or α-linolenic acid to the growth medium, since this yeast is incapable of producing these fatty acids.[82] A gene expression cassette was created containing the codon optimized Δ9 elongase (IgASE2), which was subsequently introduced into *T. oleaginosus*.

After screening of several transformants the fatty acid profile of the best producer was analyzed after 72 hours cultivation in YPD medium. The analysis revealed the production of 17% eicosadienoic acid (EDA) (TFA) and trace amounts of eicosatrienoic acid (ETE) (1% TFA) (Figure 12).

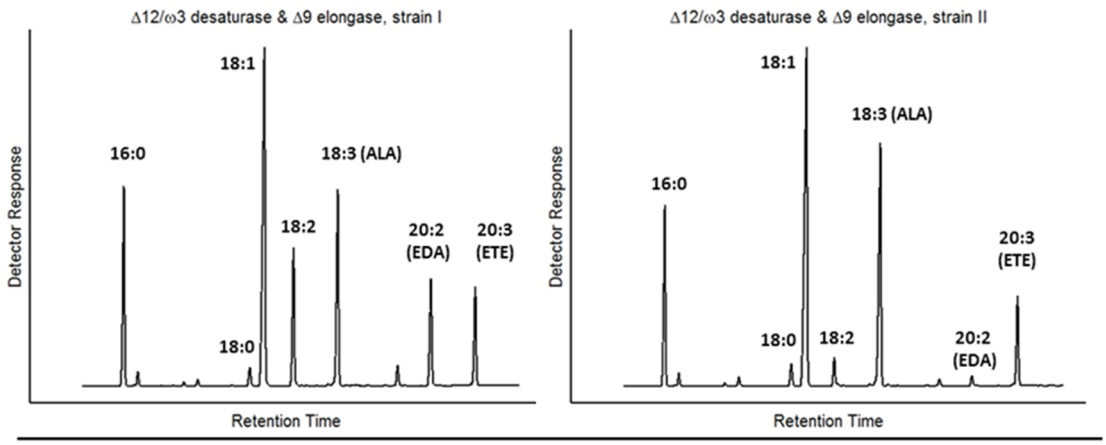


Fatty Acid [% (TFA)]	C16:0	C18:0	C18:1	C18:2	C18:3 (ALA)	C20:2 (EDA)	C20:3 (ETE)
Wild type	13.5 ± 0.7	1.7 ± 0.6	35.7 ± 0.2	46.2 ± 1.8	2.8 ± 0.5	0.0	0.0
Δ12/ω3 desaturase	18.1 ± 0.5	2.7 ± 0.2	32.3 ± 2.8	27.7 ± 2.0	1.3 ± 0.4	16.8 ±0.6	1.0 ± 0.3

Figure 12. GC-FID chromatograms and corresponding fatty acid distribution of a transgenic *T. oleaginosus* strain expressing the Δ9 elongase IgASE2 from *I. galbana*. Cultivation was done in YPD medium for three days.

IgASE2 does not discriminate between linoleic and α-linolenic acid and both substrates have been reported to be metabolized at equal shares in *S. cerevisiae*.[82] Consequently, the small amount of eicosatrienoic acid can be explained by turnover of the small α-linolenic acid content (3% TFA) observed in the wild type yeast. The amount of oleic acid was similar to the wild type and contained 32% (TFA). As expected the linoleic acid content decreased due to the metabolic turnover by the IgASE2 from 46% (TFA) to 26% (TFA). This finding depicts a yield of 56.5% and is remarkably similar to the reported turnover in *S. cerevisiae* (55.7%) using exogenous linoleic acid as substrate.[82] Consequently, it can be assumed that the reconstitution of IgASE2 in *S. cerevisiae* and *T. oleaginosus* works comparably well.

To increase the production of eicosatrienoic acid, a *T. oleaginosus* strain was engineered to express the bi-functional Δ12/ω3 fatty acid desaturase (Fm1) and the Δ9 elongase (IgASE2). Therefore, the T-DNA containing the gene expression cassettes for the hygromycin B resistance gene (*hph*) and the desaturase gene *fm1* was expanded by a gene expression cassette for the elongase *IgASE2* gene. The resulting plasmid pRF-HU2-(GPD)-Fm1-IgASE2 was transformed in *T. oleaginosus*. It was observed that the transformation efficiency was reduced compared to previous *Agrobacterium* mediated transformations resulting in lower amounts of transformants. Consequently, the transformation procedure was repeated multiple times to obtain a sufficient number of transformants that were subsequently screened for the production of eicosatrienoic acid. Interestingly, it was possible to identify two transformants with varying fatty acid profiles. After cultivation in YPD medium for 3 days, the first transformant accumulated 17% (TFA) α-linolenic acid and equal amounts (9% (TFA)) of eicosadienoic acid and eicosatrienoic acid (Figure 13).



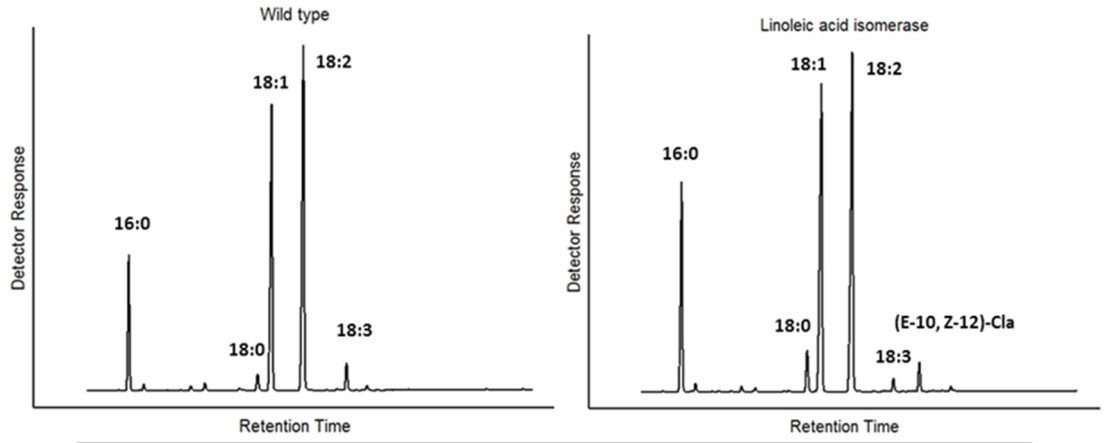
Fatty Acid [% (TFA)]	C16:0	C18:0	C18:1	C18:2	C18:3 (ALA)	C20:2 (EDA)	C20:3 (ETE)
Wild type	13.5 ± 0.7	1.7 ± 0.6	35.7 ± 0.2	46.2 ± 1.8	2.8 ± 0.5	0.0	0.0
Δ9 elongase Δ12/ω3 desaturase, strain I	15.0 ± 0.5	1.6 ± 0.1	36.3 ± 0.8	11.3 ± 1.2	17.0 ± 1.7	9.7 ± 0.4	8.9 ± 1.1
Δ9 elongase Δ12/ω3 desaturase, strain II	14.9 ± 0.2	2.3 ± 0.1	42.3 ± 1.0	2.2 ± 0.4	28.5 ± 1.5	0.9 ± 0.1	9.0 ± 0.3

Figure 13. GC-FID chromatograms and the corresponding fatty acid distribution of different phenotypes of transgenic *T. oleaginosus* strains expressing IgASE2 and Fm1. Cultivation was done in YPD medium for three days.

By contrast, the fatty acid distribution of the second transformant encompassed 28% (TFA) α-linolenic acid, 0.8% (TFA) eicosadienoic acid and 9% (TFA) eicosatrienoic acid. Despite the difference in the intracellular α-linolenic acid pool, the amount of eicosatrienoic acid was equivalent in both strains, indicating a potential degradation by β-oxidation. The significant difference in eicosadienoic acid (9% strain I; 0.8% strain II) (TFA) can be explained by different turnover of linoleic acid found in both strains. While the first strain displays a sufficient linoleic acid content (11% TFA) as metabolic precursor, the second strain only comprises 2%, thus resembling a metabolic bottleneck. The differences in both strains derived from transformation using the identical T-DNA is remarkable. It shows the significant effect caused by random genomic integration of the T-DNA and the importance of a screening method to identify beneficial transformants.

2.2.6 Engineering *T. oleaginosus* for the Production of Conjugated (E10, Z12) Linoleic Acid

The potential of *T. oleaginosus* to produce conjugated linoleic acid (CLA) was evaluated. Therefore the codon optimized linoleic acid isomerase PAI from *P. acnes* was chosen, since it has been successfully reconstituted in *Y. lipolytica*.[87] PAI is a soluble cytosolic protein, which transforms free linoleic acid (*Z*9, *Z*12) into (*E*10, *Z*12) conjugated linoleic acid.[87] Several *T. oleaginosus* transformants expressing PAI were screened, which allowed the identification of a recombinant strain that was able to accumulate 2.6% (TFA) CLA after three days cultivation in YPD medium (Figure 14).



Fatty Acid [% (TFA)]	C16:0	C18:0	C18:1	C18:2	C18:3 (ALA)	CLA
Wild type	13.5 ± 0.7	1.7 ± 0.6	35.7 ± 0.2	46.2 ± 1.8	2.8 ± 0.5	0.0
Linoleic acid isomerase	17.6 ± 0.9	3.9 ± 0.4	35.1 ± 0.6	39.5 ± 1.2	1.3 ± 0.0	2.6 ± 0.1

Figure 14. GC-FID chromatograms and corresponding fatty acid distribution of a transgenic *T. oleaginosus* strain expressing the linoleic acid isomerase PAI from *P. acnes*. Cultivation was done in YDP medium for 3 days.

The yield of CLA depends on the availability of free cytosolic fatty acid, which is not acetylated to coenzyme A. Comparable data for *Y. lipolytica* expressing PAI fatty acids showed CLA accumulation of 5.9% (TFA).[87] Since *Y. lipolytica* is well known for its high concentrations of free cytosolic fatty acids, the possible yields are generally quite low.[87] These results demonstrate that *T. oleaginosus* can accumulate sufficient free fatty acids, which can be converted by free fatty acid modifying enzymes such as PAI.

2.2.7 Influence of Carbon Source on Recombinant Lipid Productivity

The analysis of the fatty acid profiles from the engineered *T. oleaginosus* strains revealed that these transgenic strains are capable of producing large amounts of novel fatty acids in YPD medium. In the next step, this study was extended on different media compositions. As YPD medium is a rich medium and contains an abundant supply of nitrogen and phosphate, it is not applicable for the production of large amounts of lipids. Therefore a series of nutrition limited media comprising different carbon sources were applied. Lipid accumulation was triggered by either nitrogen or phosphate limitation using the C-6 sugars glucose and N-acetylglucosamine as well as the C-5 sugar xylose. For lipid accumulation by nitrogen limitation, a C/N ratio of 102 was applied in glucose and xylose containing media. In contrast, YPD medium contains a C/N ratio of 1.9. The C/P ratio for lipid accumulation by phosphate limitation in media containing N-acetylglucosamine was 389. The sugars used in the cultivations represent typical carbon sources, which are also available in waste biomass feedstocks such as cellulose (glucose), hemicellulose (xylose) hydrolysate and crab shell hydrolysate (N-acetylglucosamine). Cultivation of the transgenic *T. oleaginosus* strains was done for 7 days in triplicates using shake flasks harboring 100 mL growth media. The lipid content and the associated fatty acid distribution was monitored after 24, 72 and 168 hours and compared to the wild type *T. oleaginosus* control.

All transgenic *T. oleaginosus* showed comparable growth kinetics, biomass and lipid production in reference to the wild type. The biomass (DCW) reached between 10-12 g/L after 24 hours and gradually increased to 12- 14 g/L after 72 to 168 hours. In contrast, the lipid content varied significantly depending on the growth media applied. The lipid storage in YPD medium reached about 10% after 72 hours, but was significantly increased when growth media with N- or P-limitation was applied. When lipid accumulation was triggered by N-limitation in xylose and glucose containing media, the highest lipid contents were observed. For instance after 72 hours the lipid production peaked, yielding 57% (DCW) when xylose was used as carbon source and 48% (DCW) in the case of glucose. In contrast, when *T. oleaginosus* was grown on N-acetylglucosamine the less effective P- limitation led to a maximum lipid accumulation of 35% (DCW) after 72 hours. In general after 24 hours the lipid content in all cultivations was about half the amount produced at 72 hours. In the remaining cultivation period, the lipid content stayed constant.

Regarding the time dependent fatty acid composition, significant changes were observed between rich and limiting media in the wild type as well as the transgenic *T. oleaginosus* cells (see Appendix Table 1 for comprehensive data). Generally, an increase in the fatty acid saturation degree was observed when minimal media with N- or C-limitation was used. Particularly, the linoleic acid content changed from 46% (TFA) in rich media to 6 -8% (TFA) in lipid production media. Consequently, the drop in linoleic acid, which resembles the metabolic precursor for the production of α-linolenic acid, eicosadienoic acid, eicosatrienoic acid and conjugated linoleic acid led to a decrease in their production titers. Figure 15 shows the time dependent TFA content of α-linolenic acid (ALA), ecosadienoic acid (EDA), ecosatrienoic acid (ETE) and conjugated linoleic acid (CLA) on YPD medium as well as substrate limitation medium supplemented with glucose (Glc), xylose (Xyl) and N-acetylglucosamine (NacGlc) as carbon source.

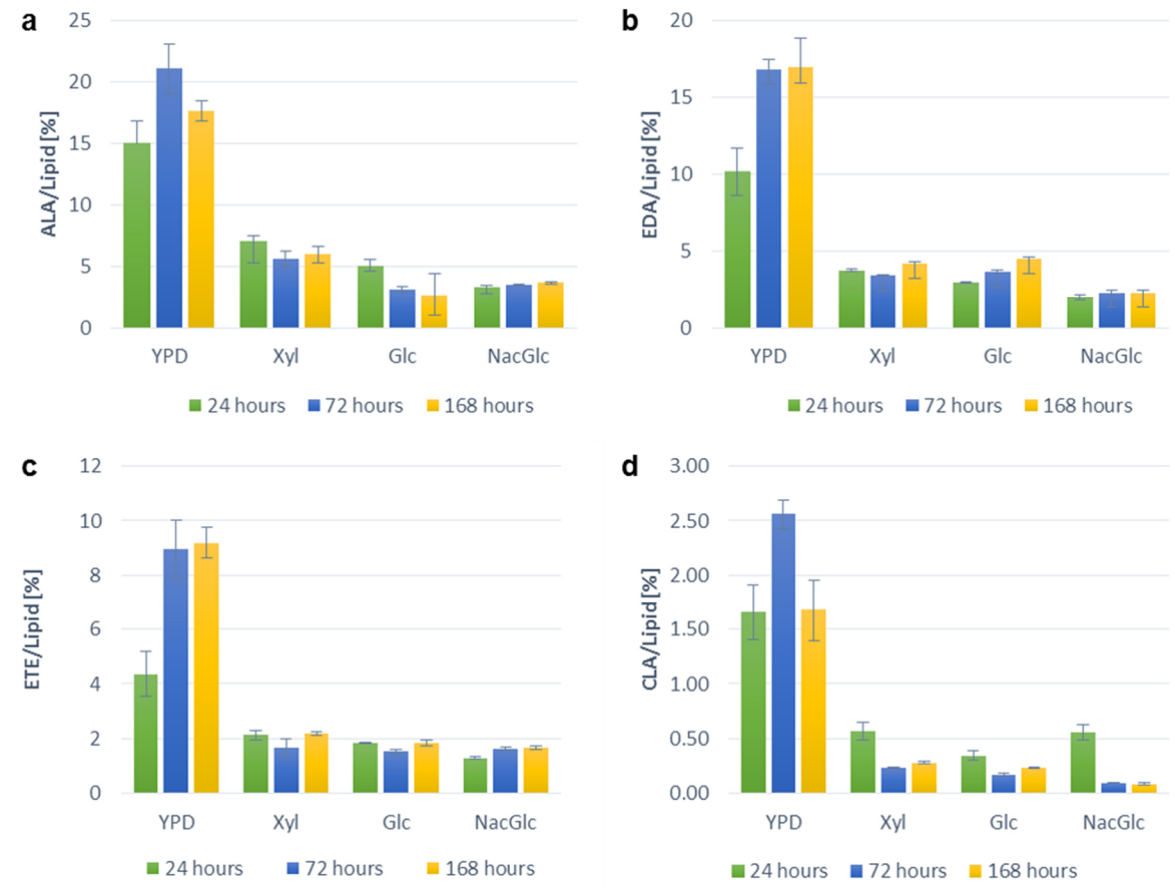


Figure 15. Time dependent total fatty acid distribution (TFA). Selected fatty acids such as α-linolenic acid (ALA) (**a**), ecosadienoic acid (EDA) (**b**), ecosatrienoic acid (ETE) (**c**) and conjugated linoleic acid (CLA) (**a**) derived from cultivations in different media compositions (YPD medium (YPD), limitation medium with xylose (Xyl), limitation medium with glucose (Glc) and limitation medium with N-acetylglucosamine (NacGlc)) are illustrated.

As expected the highest total fatty acid contents were found in YPD medium, which produced the highest linoleic acid content (TFA). As a result the total fatty acid contents of ALA, EDA, ETE and CLA were measured at 21%, 17%, 9% and 2.6%, respectively. Regarding the substrate

limitation medium the yields were much lower. The ALA yields were found between 3-7%, EDA was measured between 2-5%, ETE was accounted between 1-2% and CLA was detected to be less than 1%. The limitation medium supplemented with xylose as carbon source showed slightly increased production yields compared to glucose and N-acetylglucosamine containing growth media. This finding is in good agreement with the elevated linoleic acid content, which was observed in the wild type strain (8% (TFA)) compared to glucose (6% (TFA)) and N-acetylglucosamine (7% (TFA)).

Interestingly, the total fatty acid content (TFA) of ALA and CLA from YPD cultivations peaked at 72 hours and decreased towards the end of the cultivation (168h), while the content of EDA and ETE remained stable. Consequently, in YPD medium only EDA and ETE seem to be metabolically stable, while ALA and CLA are presumably subject to degradation. However, in limitation medium only the CLA production decreases significantly after 24 hours, leaving the total fatty acid content of ALA, EDA and ETE relatively stable during the entire cultivation period.

When the total fatty acid contents of ALA, EDA, ETE and CLA are referred to the total lipid contents per dry cell weight (DCW) a different picture appears (Figure 16) (see Appendix Table 2 for comprehensive data). The highest share of the engineered fatty acids was observed in YPD medium, but in this medium *T. oleaginosus* strain only produces about 10% lipids (DCW). In contrast, the lipid content is significantly increased in limitation medium, but resulted in much lower shares of the engineered fatty acids. Figure 16 shows the ALA, EDA, ETE and CLA contents in percentage of the dry cell weight.

Surprisingly, the yields derived from YPD and limitation medium are very comparable. The highest titres for ALA, EDA, ETE and CLA were found at 3.4%, 2.4%, 1.2% and 0.16% respectively. While the highest yields for ALA, EDA and ETE were found in xylose limitation medium after 168 hours, the CLA content peaked in YPD medium after 72 hours. In general, N-acetylglucosamine limitation medium showed the lowest production yields in all production strains, which might result from phosphate starvation.

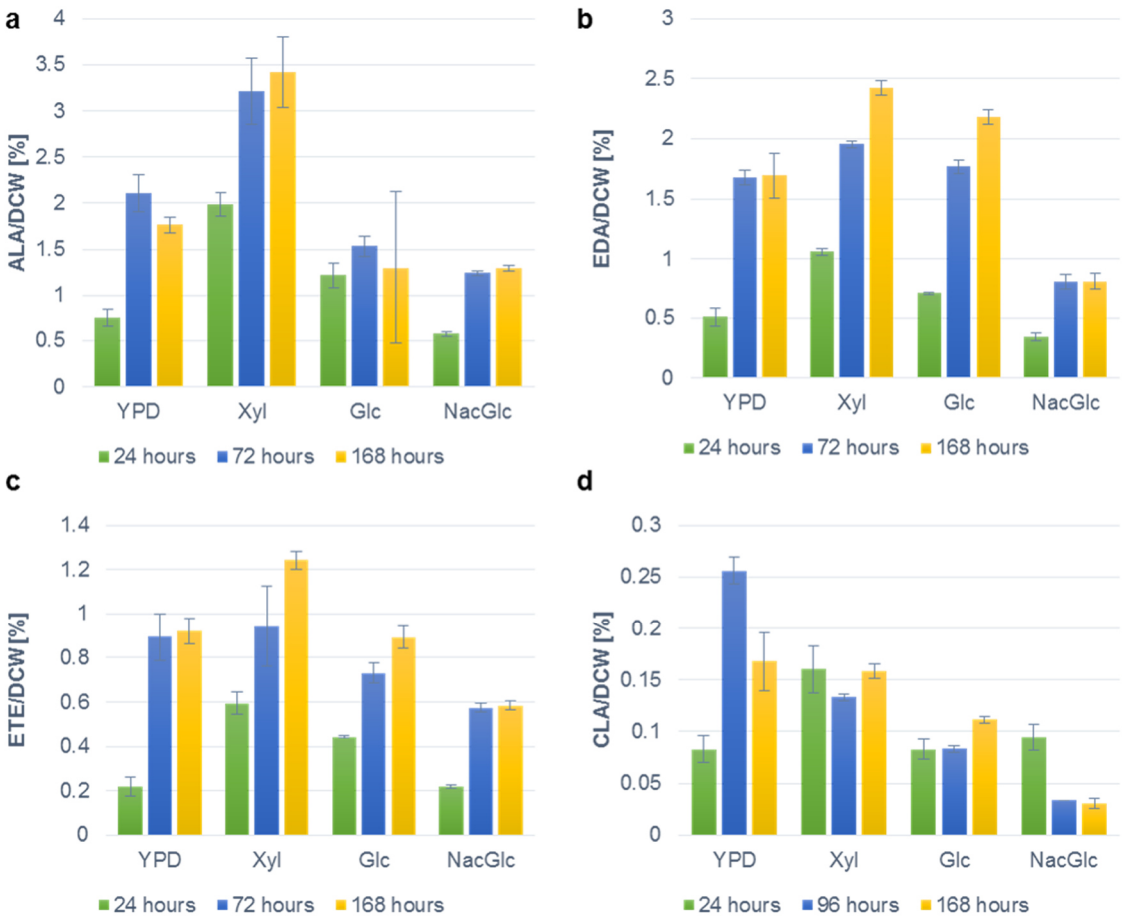


Figure 16. Time dependent total fatty acid accumulation as share of the dry cell weight. Selected fatty acids such as α-linolenic acid (ALA) (**a**), ecosadienoic acid (EDA) (**b**), ecosatrienoic acid (ETE) (**c**) and conjugated linoleic acid (CLA) (**a**) derived from cultivations in different media compositions (YPD medium (YPD), limitation medium with xylose (Xyl), limitation medium with glucose (Glc) and limitation medium with N-acetylglucosamine (NacGlc)) are illustrated.

The biomass production is proportional to the amount of nutrition and in particular the amount of the carbon source available in the fermentation medium. The limitation media used for the shake flask experiments contained 30 g/L of the corresponding sugars, while the rich YPD medium contained 20 g/L glucose. These values translated into an average biomass formation of 13 g/L (DCW) and result in the production of ALA, EDA, ETE and CLA with the highest yields of 0.44 g/L, 0.38 g/L, 0.17 g/L and 0.02 g/L, respectively. An increase in the carbon content would translate into enhanced biomass production and would ultimately lead to a higher lipid production. For example *T. oleaginosus* has been reported to grow in a bioreactor to a cell density of 118 g/L yielding 30 g/L lipids.[69] In contrast, increasing the cell density in shake flask cultivations quickly leads to insufficient oxygen intake and a retarded metabolism. Hence, the amount of carbon in the shake flask cultivations was not increased further. Presumably, the production yields of the engineered fatty acids can be significantly increased by cultivation optimizations and the application of bioreactors.

2.2.8 Influence of Carbon Source on the GDP Promotor Strength

The YPD cultivations have demonstrated the feasibility of *T. oleaginosus* to produce high amounts of ALA, EDA, ETE and CLA as total fatty acid content. However, when the medium was changed to promote lipid accumulation, the yields dropped significantly. The production levels of the engineered fatty acids depend on the catalytic efficiency of the corresponding fatty acid modifying enzyme, its promotor strength and the metabolic stability of the produced fatty acid. Since the enzyme was functionally expressed and displayed sufficient catalytic efficiency in the YPD cultivations, the strength of the GPD promotor was analyzed in the next step. The GPD promotor was initially chosen, since it displays a strong constitutive expression level. As part of the glycolysis it is furthermore involved in metabolism of glycose and N-acetylglucosamine and should remain active during lipogenesis. Moreover, the GDP promotor has been successfully used for the expression of the Δ12 desaturase and Δ9 elongase the oleaginous yeast *Y. lipolytica*.[5] However, this data was deduced from a two-step cultivation comprising a 2 day cultivation period in rich medium followed by 5 days cultivation in a high glucose and nitrogen free medium.[5] Even though a direct comparison is not possible, the used GDP promotor in *Y. lipolytica* was feasible for the production of high amounts of eicosapentaenoic acid.

To test the strength of the GDP promotor in different cultivation media, the YFP reporter protein under the control of the GPD promotor was used. The expression level was implied as fluorescence signal, which was quantified in rich and lipid production media using the same cultivation condition as before. The results are illustrated in figure 17.

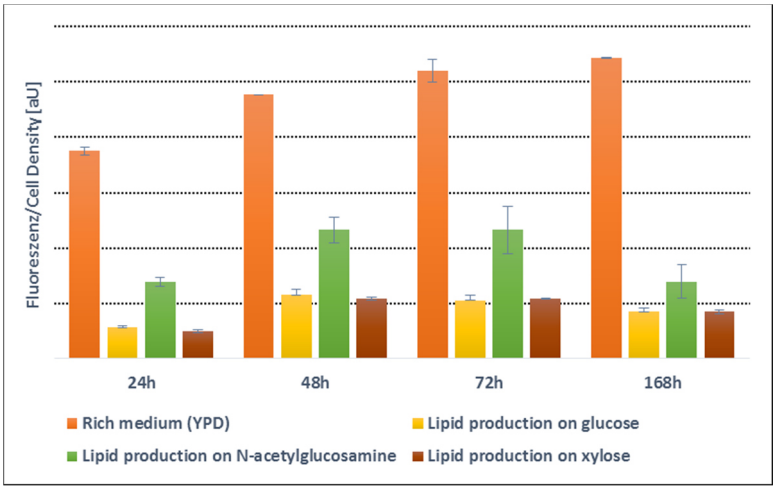


Figure 17. Influence of the growth media on YFP expression using the GPD promotor.

As expected, the highest signal was found in YPD rich medium, which ultimately led to the highest production titers of modified fatty acids. The fluorescence level was reduced at about two third when cultivation was taking place in phosphate limitation medium with N-acetylglucosamine as carbon source. Interestingly, the signal further decreased to one fifth when limitation media with glucose or xylose as carbon source was applied.

The data indicates that the GPD promotor strength is upregulated in medium with high nitrogen content (YPD medium) and down regulated in limitation media. Consequently, these data clearly show a cultivation media effect on the promotor strength, which reflects the observed production titers in particular from rich to minimal media. Accordingly, the promotor seems to be a good choice for heterologous expression in nitrogen and carbon rich medium. But a decrease in strength in lipid production media makes it reasonable to search for promotors that are independent of the cultivation medium used or upregulated in limiting media.

The metabolic stability of the engineered fatty acids is experimentally difficult to address and was not done in this study. However, the analysis of the promotor strength revealed an increase in the YFP expression in N-acetylglucosamine containing medium, which unexpectedly did not lead to a superior production yield compared to glucose and xylose. Due to this finding, it cannot be ruled out that the engineered fatty acids are subject to degradation during lipid accumulation. The fatty acid content of oleaginous yeasts is generally characterized by a high oleic acid and a low linoleic acid content when cultivated under lipid production conditions. Consequently, during lipogenesis unsaturated fatty acids such as linoleic acid but also ALA, EDA, ETE and CLA might be degraded to acetyl-CoA and rebuilt as oleic acid. To prevent fatty acid degradation β-oxidation knockout mutants analog to *pex10* or *mfe1* in *Y. lipolytica* should be created, which were demonstrated to increase the lipid content.[5,124]

Alternatively, the decrease of linoleic acid may result in an excessive production of oleic acid, which ultimately leads to an attenuation of the linoleic acid content. In this scenario, heterologous expression of the bi-functional enzyme Fm1 is insufficient to compensate for the large decrease of linoleic acid. As Fm1 contains a Δ12 desaturase activity to convert oleic acid into linoleic acid its metabolic levels should be replenished. However, this anticipated effect is not seen in the cell cultures. Therefore, homologous overexpression of the *T. oleaginosus* Δ12 desaturase should result in increased linoleic acid levels that serves as platform for the production of very long chain polyunsaturated fatty acids.

2.3 Conclusion and Outlook

In this study an *A. tumefaciens* mediated transformation (ATMT) protocol for the oleaginous yeast *T. oleaginosus* was established. Transgenic yeasts were successfully selected using the dominant selection marker hygromycin B. A 800 bp long fragment of the GPD promotor in connection with a 600 bp long GDP terminator facilitated the strong heterologous expression of a YFP reporter protein. Furthermore, it was demonstrated that *T. oleaginosus* can be engineered to produce significant concentrations of high value non-native fatty acids such as ALA, EDA, ETE and CLA. It was found that the total fatty acid content of the engineered fatty acids strongly depends on the media applied. In rich medium the relative yields of the respective fatty acids were significantly increased compared to nutrient deficient media. In respect to the total intracellular lipid content, the highest content of engineered fatty acids was found in xylose supplemented medium using N-limiting conditions. Analysis of the GDP promotor strength in rich and nutrient deficient media showed a clear decrease of the expression levels when substrate limitation was applied. Hence, it makes only sense to use the GPD promotor in rich medium, which is in contrast to the oleaginous yeast *Y. Lipolytica*. To increase the production of high value fatty acids in *T. oleaginosus* a variety of strategies could be applied. Apart from creating β-oxidation knockout mutants, the overexpression of the diacylglycerol acyltransferase (DGAT), malic enzyme (ME), and AMP deaminase (AMPD) has been shown to facilitate lipid production in *Y. Lipolytica*.[124] Furthermore, leucine has been suggested to be an essential trigger for the lipogenesis and might be explored to increase the lipid content.[124] In this study data on lipid production and the corresponding fatty acid composition was collected from limited shake flask cultures. Optimization of the growth conditions in a bioreactor are expected to increase production yields.

2.4 Part I: Materials and Methods

Yeast and Bacterial Strains

The oleaginous yeast *Trichosporon oleaginosus* ATCC 20509 (DSM-11815) was obtained from the Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ) (Braunschweig, Germany).

Agrobacterium tumefaciens strain AGL1 (BAA-101) used for the transformation of *Trichosporon oleaginosus* was acquired from the American Type Culture Collection (ATCC) (Manassas, VA, USA). *Escherichia coli* strain XL1-Blue (recA1, endA1, gyrA96, thi-1, hsdR17, supE44, relA1, lac[F′proAB lacI^{q}ZΔM15Tn10(Tet^{r})]) was used for cloning and was obtained from Novagen/Merck Millipore (Schwalbach, Germany).

Growth Media

Trichosporon oleaginosus was routinely cultivated in YPD medium (glucose, 20 g/L; tryptone, 20 g/L; yeast extract, 10 g/L). For the lipid accumulation nitrogen limitation medium (glucose or xylose, 30 g/L; yeast extract, 0.5 g/L; $(NH_4)_2SO_4$, 0.3 g/L; $MgSO_4 \cdot 7H_2O$, 1.5 g/L; KH_2PO_4, 2.4 g/L; Na_2HPO_4 0.91 g/L; $CaCl_2 \cdot H_2O$, 0.22 g/L; $ZnSO_4 \cdot 7H_2O$, 0.55 µg/L; $MnCl_2 \cdot 4H_2O$, 22.4 µg/L; $CuSO_4 \cdot 5H_2O$, 25 µg/L; $FeSO_4 \cdot 7H_2O$, 25 µg/L, pH 6.1) or phosphate limitation medium N-acetylglucosamine, 30 g/L; yeast extract, 0.5 g/L; NH_4Cl, 0.5 g/L; $MgSO_4 \cdot 7H_2O$, 1.5 g/L; KH_2PO_4, 0.11 g/L; Na_2HPO_4, 38.7 mg/L; $CaCl_2 \cdot 2H_2O$, 0.22 g/L; $ZnSO_4 \cdot 7H_2O$, 0.55 µg/L; $MnCl_2 \cdot 4H_2O$, 22.4 µg /L; $CuSO_4 \cdot 5H_2O$, 25 µg/L; $FeSO_4 \cdot 7H_2O$, 25 µg/L) was used.

E.coli XL1-Blue and *Agrobacterium tumefaciens* AGL1 was cultivated in LB medium (tryptone, 10 g/L; yeast extract, 5 g/L; NaCl, 10 g/L)

Media used for the *Agrobacterium tumefaciens* mediated transformation of *Trichosporon oleaginosus* were liquid induction medium with acetosyringone (L-IMAS) (K_2HPO_4, 2.05 g/L; KH_2PO_4, 1.45 g/L; NaCl, 0.15 g/L; $MgSO_4 \cdot 7H_2O$, 0.5 g/L; $CaCl_2 \cdot 2H_2O$, 67.0 mg/L; 2- 4-Morpholineethanesulfonic acid monohydrate (MES), 7.8 g/L; glucose, 1.8 g/L; acetosyringone, 39.24 mg/L; $FeSO_4 \cdot 7H_2O$, 2.5 mg/L; $(NH_4)_2SO_4$, 0.5 g/L; glycerol 5% (*v*/*v*); trace elements solution (100 mg Na_2MoO_4, $MnSO_4 \cdot H_2O$, $ZnSO_4 \cdot 7H_2O$, $CuSO_4 \cdot 5H_2O$, H_3BO_3 in 1L dd$H_2$0, 5% (*v*/*v*), pH 5.6) and solid induction medium with acetosyringone (S-IMAS) (equivalent to L-IMAS, without glucose and supplemented with 18 g/L agar).

Cloning

General Methods

Cloning was done using the restriction enzyme free cloning technique "USER Cloning" from New England Biolabs (Ipswich, MA, USA). Polymerase chain reactions were performed using the PfuTurbo Cx Hotstart DNA Polymerase from Agilent Technologies (Santa Claram, CA, USA). PCR primers were synthesized by Eurofins MWG Operon. The reactions were done in 50 µL volumes using 10 ng template DNA, 10 pmol phosphorylated primers (Eurofins MWG Operon), 0.2 mM of each deoxynucleotide triphosphate (dNTPs) and 1.25 U PfuTurbo Cx Hotstart DNA Polymerase. PCR thermocycles were performed based on the applied oligonucleotides and according to manufacturer's recommendation of the PfuTurbo Cx Hotstart DNA Polymerase. PCR products were purified by a 1% (*v/v*) agarose gel. Gene bands were cut out of the agarose gel and were extracted using the innuPREP DOUBLEpure Kit (Analytik Jena, Jena) according to manufacturer's protocol. Purification of digested DNA by restriction or nicking enzymes was done by using the innuPREP DOUBLEpure Kit (Analytik Jena, Jena) using manufacturer's protocol. Plasmids from *E.coli* XL1-Blue were extracted and purified using the GeneJET Plasmid Miniprep Kit (Thermo Scientific, Braunschweig) according to manufacturer's protocol. Chemical competent *E.coli* as well as transformation procedures were followed as described elsewhere.[125] All clones were validated by DNA sequencing.

DNA Templates

The plasmid pRF-HU2 obtained from the Fungal Genetics Stock Center (FGSC) (Manhattan, KS, USA) contained the T-DNA for the *A. tumefaciens* mediated transformation and was used as template to create the plasmids pRF-HU2(GPD)-rDNA, pRF-HU2(GPD), pRF-HU2-(GPD)-YFP, pRF-HU2-(GPD)-IgASE2, pRF-HU2-(GPD)-Fm1, pRF-HU2-(GPD)-PAI and pRF-HU2-(GPD)-IgASE2-Fm1.

Genes were synthesized by Life Technologies (Carlsbad, CA, USA) and are illustrated in the Appendix Figure 18-21. Genes for YFP reporter protein, Δ9 elongase IgASE2 from *Isochrysis galbana*, Δ12/ω3 desaturase Fm1 from *Fusarium moniliforme* and linoleic acid isomerase PAI from *Propionibacterium acnes* were codon optimized based on the preferred codon usage table for the glyceraldehyde-3-phosphate dehydrogenase (GPD) (Genebank AF126158.1). Codon usage was analyzed by the "Sequence Manipulation Suite" using the program "Codon Usage". Codon optimized genes were designed by "Sequence Manipulation Suite" using the program "Reverse Translate".[126] All genes were synthesized as cassettes containing 800 bp of the GPD

promotor and 600 bp of the GPD terminator derived from the NCBI genome JZUH00000000.1. Genomic DNA of *T. oleaginosus*, which was isolated using the Yeast DNA Extraction Reagent Kit from Thermo Scientific (Waltham, MA, USA) according to the manufacturer procedure.

Cloning of the Plasmid pRF-HU2(GPD)-rDNA

A 1 kb of the 18S rDNA and 1.1 kb of the ITS/26S rDNA of *Trichosporon oleaginosus* were amplified by PCR. Primers used for the 18S rDNA segment were 5′-GGT CTT AAU CCA GTA GTC ATA TGC TTG TCT C-3′ and 5′-GGC ATT AAU CCT AGT CGG CAT AGT TTA C-3′. Primers for the ITS/26S rDNA segment were 5′-GGA CTT AAU TGA ACC TGC GGA AGG ATC AT-3′ and 5′-GGG TTT AAU CAT CCT AAG CTC GAA CGT GTC C-3′. The annealing temperature for both primer pairs was 60°C. Both DNA fragments were ligated into the pRF-HU2 plasmid, which was digested by restriction enzymes PacI and Nt.BbvCI (New England biolabs). USER friendly cloning was done as described by Frandsen et al.[101] and yielded the plasmid pRF-HU2-rDNA. In the next step, the tryptophan promotor (*Aspergillus nidulans*) from the hygromycin B resistance cassette in the plasmid pRF-HU2-rDNA was exchanged by a 390 bp long fragment from the glyceraldehyde-3-phosphate dehydrogenase (GPD) promotor of *T. oleaginosus*. Therefore the plasmid pRF-HU2-rDNA was linearized by PCR excising the tryptophan promotor. Primers used were 5′-ATG AAA AAG CCU GAA CTC ACC GCG-3′ and 5′-ATT AAT GCC UCA GCG AAC CTG CGG AAG GAT-3′. The annealing temperature was 55°C. The 390 bp-GPD DNA fragment was amplified of the *T. oleaginosus* genomic DNA using the primers 5′-AGG CTT TTT CAU TGT TGA TCA AGT TGA TTT TTG GG-3′ and 5′-AGG CAT TAA UCC TCC TCC GGC ACC-3′ with the annealing temperature of 65°C. USER friendly ligation of both DNA fragments yielded the plasmid pRF-HU2(GPD)-rDNA. USER ligation was done following the manufactures protocol.

Cloning of the Plasmid pRF-HU2(GPD)

The pRF-HU2(GPD) unlike the pRF-HU2(GPD)-rDNA was missing the flanked rDNA sequences for homologous recombination in the rDNA of *T. oleaginosus*. To create the plasmid pRF-HU2(GPD) the tryptophan promotor from the hygromycin B resistance cassette in the plasmid pRF-HU2 was exchanged with the 390 bp-GPD DNA fragment of *T. oleaginosus*. Primers used for the PCR reaction were 5′-AGG CTT TTT CAU TGT TGA TCA AGT TGA TTT TTG GG-3′ and 5′-AGG CAT TAA UCC TCC TCC GGC ACC-3′ in combination with the annealing temperature of 58°C. The pRF-HU2 plasmid was linearized by PCR using the primers 5′-ATG AAA AAG CCU GAA CTC ACC GCG-3′ and 5′-ATT AAT GCC UAT CGA TGG GCC CGC TGA G-3′. The annealing

temperature was 65°C. Ligation was carried out following USER Cloning standard protocols and yielded the plasmid pRF-HU2-(GPD).

Cloning of the Plasmids pRF-HU2-(GPD)-YFP, pRF-HU2-(GPD)-IgASE2, pRF-HU2-(GPD)-Fm1 and pRF-HU2-(GPD)-PAI.

For the heterologous gene expression, pRF-HU2-(GPD) was extended by another gene cassette. These cassettes were created using codon optimized genes of either the YFP reporter protein, the Δ9 elongase IgASE2 from *Isochrysis galbana*, the Δ12/ω3 desaturase Fm1 from *Fusarium moniliforme* or the linoleic acid isomerase PAI from *Propionibacterium acnes*. The genes were synthesized as cassettes containing 800 bp of the GPD promotor and 600 bp of the GPD terminator derived from the NCBI genome JZUH00000000.1. The expression cassettes were integrated into the pRF-HU2-(GPD) in the same orientation as the hygromycin B resistance cassette on its 3′ end. Cloning was performed in analogy to the creation of the pRF-HU2-(GPD) plasmid. Therefore the pRF-HU2-(GPD) plasmid was linearized by PCR using the primers 5′-ATT AAA CCC UAT GCC TCA GCA CTA GTC-3′ and 5′-ATT AAG ACC UGA CCT CAG CAA GCT TCG TGA C-3′ (55°C annealing temperature). The diverse gene expression cassettes were amplified using the primers 5′- AGG TCT TAA UAT CCG CTG ACA TTG GAC CTT-3′ and 5′-AGG GTT TAA UGG GGA TTG GCG TCA TCA AGT GC -3′ (60°C annealing temperature). USER ligation of the PCR products from the pRF-HU2-(GPD) and cassettes yielded the plasmids pRF-HU2-(GPD)-YFP, pRF-HU2-(GPD)-IgASE2, pRF-HU2-(GPD)-Fm1 and pRF-HU2-(GPD)-PAI.

Cloning of the Plasmid pRF-HU2-(GPD)-IgASE2-Fm1

For the production of ecosatrienoic acid (ETE) the plasmid pRF-HU2-(GPD)-IgASE2 was extended by the Fm1 expression cassette, which led to the creation of the plasmid pRF-HU2-(GPD)-IgASE2-Fm1. The Fm1 cassette was integrated using the same orientation, subsequent to the 3′ end of the IgASE2 cassette. The Fm1 cassette was introduced into the pRF-HU2-(GPD)-IgASE2 plasmid using the HindIII restriction site. Therefore the Fm1 cassette was flanked with a HindIII restriction site using the primers 5′-ATA TAA GCT TGG GGA TTG GCG TCA TCA AGT-3′ and 5′-ATT AAA GCT TAT CCG CTG ACA TTG GAC CTT TTG G-3′ (65°C annealing temperature). The resulting PCR product as well as the pRF-HU2-(GPD)-IgASE2 plasmid was digested with the HindIII restriction enzyme (Thermo Scientific, Braunschweig) according to manufacturer's recommendation. Ligation of both DNA fragments was done using the T4 ligase (Thermo Scientific, Braunschweig) according to the manufacturer's protocol. Plasmids obtained by ligation were analyzed for the correct orientation of the FM1 expression cassette. Therefore

a PCR was performed using the primers IgASE2_F: 5′-CGA GAA GAA GGG CGC CTA CC-3′ Fm1_R: 5′-TTG AGG GTG CCG AAG ATG GT-3′ (60°C annealing temperature). Only clones displaying the corrected DNA fragment of 3249 bp fragment were selected and confirmed by sequencing.

Transformation of *Trichosporon oleaginosus* by Electroporation and LiAc/SS Carrier DNA/PEG Method.

Preparation of Linear DNA

For the transformation by electroporation and the LiAc/SS carrier DNA/PEG method the hygromycin B resistance cassette flaked with the rDNA was amplified by PCR using the plasmid pRF-HU2(GPD)-rDNA as template. Primers used for the PCR reaction were 5′-CCA GTA GTC ATA TGC TTG TCT C-3′ and 5′-CAT CCT AAG CTC GAA CGT GTC C-3′. The annealing temperature was 55°C.

Electroporation

A Biorad MicroPulser electroporation apparatus was used. Electroporation method was applied from Thompson et al. Therefore *Trichosporon oleaginosus* from an overnight culture (YPD media 28°C) was grown in a 100 mL YPD medium (28°C) shake flask culture to an OD_{600} 1.5. The cells were washed twice with sterile double distilled water, subsequently suspended in 25 mL Tris buffer pH 7.5 containing 10 mM DTT and 0,1M LiAc and1 mM EDTA and incubated for 1 hour at 28°C with gentle agitation. Next, the cells were washed once with 25 mL ice cold double distilled water, pelleted and resuspended in 10 mL ice cold 1M sorbitol. The cells were pelleted and resuspended in ice cold 1M sorbitol to yield 500 µL. 5 µL of DNA in double distilled water were added to 40 µL cell solution and incubated on ice for 5 minutes. The mixture was transferred to a 0.2 cm electroporation cuvette, pulsed and immediately 1 mL YPD medium was added, followed by an incubation of 3 hours at 28°C. The cells (100 µL) were plated on YPD agar supplemented with 100 µM Hygromycin B. The plates were incubated for 3 -4 days at 28°C.

LiAc/SS Carrier DNA/PEG Method

Transformation was done according to Gietz et al. *Trichosporon oleaginosus* from an overnight culture (YPD medium 28°C) was grown in a 250 mL YPD medium shake flask culture to an OD_{600} 0.5 at 28°C. The cells were washed twice in 25 mL double distilled water, pelleted and resuspended in 1 mL double distilled water. 100 µL was mixed with 360 µL composed of 240

µL PEG 3350 (50% (*w/v*)), 36 µL 1M LiAc, 50 µL single stranded carrier DNA (Herring Sperm DNA, Promega) and 34 µL purified DNA from PCR reaction. The cells were resuspended by vortexing. In the next step the mixture was incubated at 42°C for 40 minutes. The cells were pelleted and resuspended in 1 mL YPD medium followed by an incubation of 3 hours at 28°C. 10 uL of the cell suspension was plated on YPD agar supplemented with 100 µM Hygromycin B. The plates were incubated for 3 -4 days at 28°C.

Agrobacterium tumefaciens Mediated Transformation of *Trichosporon oleaginosus*

The plasmids pRF-HU2-(GPD)-YFP, pRF-HU2-(GPD)-IgASE2, pRF-HU2-(GPD)-Fm1, pRF-HU2-(GPD)-PAI and pRF-HU2-(GPD)-IgASE2-Fm1 were introduced into *A. tumefaciens* using a method described elsewhere.[127] An overnight culture (LB medium supplemented with 30 µg/mL kanamycin at 28°C) was used to inoculate (OD_{600}) a 10 mL shake flask culture with L-IMAS medium and cultivated at 28°C for 6 hours. An overnight culture of *T. oleaginosus* grown in YPD medium at 28°C (OD_{600} 2-3) was centrifuged (10,000 g) and re-suspended in an appropriate volume of L-IMAS medium to dilute the cell concentration to an OD_{600} of 0.5. 500 µL of the *T. oleaginosus* and *A. tumefaciens* cell solution in L-IMAS medium were mixed. In the next step 100 µl of the *A. tumefaciens* and *T. oleaginosus* cell mixture were plated on top of an Amersham Hybond-N^+ blotting membrane from GE Healthcare (Little Chalfont, Buckinghamshire, UK) that was placed on S-IMAS agar plates. The plates were incubated at 24°C for 48 hours. Afterwards the membrane was transferred to YPD agar plates supplemented with 200 µg/mL hygromycin B and 300 µg/mL cefotaxime. Incubation of the agar plates was carried out over 5 days at 28 C.

Selection of Clones

Yeast colonies that appeared on the Hybond-N^+ blotting membrane were picked and cultivated in 5 mL YPD medium at 28°C supplemented with 300 µg/mL cefotaxime.

To find the best YFP producer 30 colonies were cultivated for 2 days. The cells were pelleted, washed with double distilled water (ddH_2O) and resuspended in ddH_2O. The cell solution was transferred to MaxiSorp F96 plates from Thermo Scientific (Waltham, MA, USA) and analyzed on a plate reader EnSpire 2 from Perkin Elmer (Waltham, MA, USA). Fluorescence intensity was measured at 527 nm (excitation 490 nm).

For the production of modified fatty acids by the plasmids pRF-HU2-(GPD)-YFP, pRF-HU2-(GPD)-IgASE2, pRF-HU2-(GPD)-Fm1 and pRF-HU2-(GPD)-PAI, 30 colonies were cultivated and

screened. For the production of ETE (pRF-HU2-(GPD)-IgASE2-Fm1) 100 colonies were cultivated and screened. Cultivation was done for 3 days in YPD medium at 28°C. After that the cells were pelleted and flash frozen for further analysis.

Cultivations to determine the lipid and fatty acid composition

To analyze the lipid content and fatty acid distribution yeast strain cultivation was carried out in 500 mL baffled shake flasks at 28°C for 7 days in triplicate. 100 mL cultivation media was used consisting either of YPD, nitrogen limitation medium with glucose/ xylose or phosphate limitation medium with N-acetylglucosamine. Cultivation was started by inoculation from an overnight culture (YPD medium supplemented with 300 µg/mL cefotaxime) at OD_{600} 0.5 in YPD medium. 15 mL samples were taken after 24, 72 and 168 hours to determine the dried cell mass, lipid content and fatty acid composition.

Determination of the DNA Concentration

DNA concentration in nuclease free water was measured by UV/VIS spectroscopy. Absorbance at 260 nm was measured for the DNA concentration. Quantification was done using the following correlation. Absorbance of A_{260} =1 corresponds to a concentration of 50 ng/µL double stranded DNA. DNA purity was determined by measuring the absorbance at 280 nm and calculating the ratio $A_{260}/A_{280.}$ Ratios between 1.8 and 2.0 indicated high purity.

Cell Mass Yield and Lipid Yield

Cellular dry weight was determined by centrifugation of the cells at 12,000 g for 10 min. The cell pellet was washed with ddH_2O and dried at 60°C to constant weight. The cellular total lipid was determined by extraction with chloroform and methanol according to the protocol of Folch et al.[128]

Fatty Acid Composition Analysis

For the fatty acid analysis, 2 ml cultivation medium was taken from shake flask cultures, washed with ddH_2O and pelleted. The wet biomass was directly converted into fatty acid methyl esters (FAME) by methanol transesterification according to the protocol of Griffiths et al.[129] FAMEs were analyzed on a GC-2010 Plus gas chromatograph from Shimadzu (Nakagyo-ku, Kyōto, Japan) with flame ionisation detector. Therefore, 1 µl sample was applied by a AOC-20i auto injector (Shimadzu) onto a ZB-WAX column (30 m, 0.32 mm ID; 0.25 µm df; Phenomenex (Torrance, CA, USA)). The initial column temperature was 150°C (maintained for 1 min). A temperature gradient was applied from 150°C – 240°C (5 °C/min), followed by 6 min

maintenance at 240°C. Fatty acids were identified according to retention times of authentic standards.

Fluorescence Microscopy

Microscopic photographs were taken on an Axio Lab. A1, fluorescence microscope equipped with an Axio Cam ICm1 (Zeiss, Oberkochen, Germany). For sample preparation the yeast cells were washed with ddH_2O and stained with Nile Red. Staining was conducted by addition of 25 µL DMSO and 25 µL (0.1 mg/mL in DMSO) of a Nile Red solution to a cell suspension. The mixture was vortexed and incubated for 10 minutes. The yeast cells were washed with ddH_2O twice and subsequently analyzed by fluorescence microscopy.

3. Part II: Targeted Engineering of cyclooctat-9-en-7-ol Synthase CotB2

Peer-Reviewed Publication

This chapter is based on the following articles:

1. "Targeted Engineering of Cyclooctat-9-en-7-ol Synthase: A Stereospecific Access to Two New Non-natural Fusicoccane-Type Diterpenes" Görner, C.; Häuslein, I.; Schrepfer, P.; Eisenreich, W.; Brück, T. *ChemCatChem* **2013**, *5*, 3289–3298.
2. "The first structure of a bacterial diterpene cyclase: CotB2." Janke, R.; Görner, C.; Hirte, M.; Brück, T.; Loll, B. *Acta Crystallogr. D. Biol. Crystallogr.* **2014**, *70*, 1528–1537.
3. "Stereoselective chemo-enzymatic oxidation routes for (1R,3E,7E,11S,12S)-3,7,18-dolabellatriene" Görner, C.; Hirte, M.; Huber, S.; Schrepfer, P.; Brück, T. *Front. Microbiol.* **2015**, *6*, 1115. Note: Only the microbial production procedure for (1R,3E,7E,11S,12S)-3,7,18-dolabellatriene from this publication is part of this thesis.

Contributions

1. CG, IH and TB conceived the study, planned and supervised experiments. CG and IH conducted metabolic engineering of *E.coli*, enzyme mutagenesis experiments and production and purification of diterpene products. PS conducted *in silico* studies. CG, IH and WE conducted NMR data collection and solved structures. CG, PS, WE and TB created the manuscript.
2. CG, RJ, TB, BL conceived the study, planned and supervised experiments. RJ and BL crystalized CotB2. CG and MH conducted metabolic engineering of *E.coli*, enzyme mutagenesis experiments and production and purification of diterpene products. CG and MH conducted NMR data collection and solved structures of fermentation products. CG RJ, MH, TB and BL created the manuscript.
3. CG, MH and TB conceived the study, planned and supervised experiments. SH helped carrying out the experiments. CG and MH conducted NMR data collection and solved structures. CG, MH, PS and TB created the manuscript.

3.1 Introduction

3.1.1 Terpenes

With over 60 000 known structures, terpenes are structurally and functionally the most diverse class of natural products.[7,8] Terpenes are derived from branched C5 carbon isoprene units forming the core carbon skeleton.[130] They are classified according to their number of carbon atoms in hemiterpenes (C5), monoterpenes (C10), sequiterpenes (C15), diterpenes (C20) triterpenes (C30) and tetraterpenoids (C40).[130] Terpenes comprise a wide range of biological functions.[32] They are used as hormones (steroids, gibberellins and abscisic acid), maintain membrane fluidity (steroids), are needed in the respiration chain (quinones) and are involved in photosynthetic light harvesting (carotenoids).[32] Prenyl group terpenes are posttranslationally added to proteins and are used for targeting and regulation.[32,131] Terpenes are produced as secondary metabolites and involved in defence and communication.[32] These secondary metabolites can be found in plants, fungi, bacteria and invertebrates displaying a series of biological effects including anti-tumor, anti-inflammatory, antiviral, antimalarial, antibiotic, fungidal and insectidal activities.[9] Furthermore, terpenes are used in plants for pollination attraction, dispersion and development.[32]

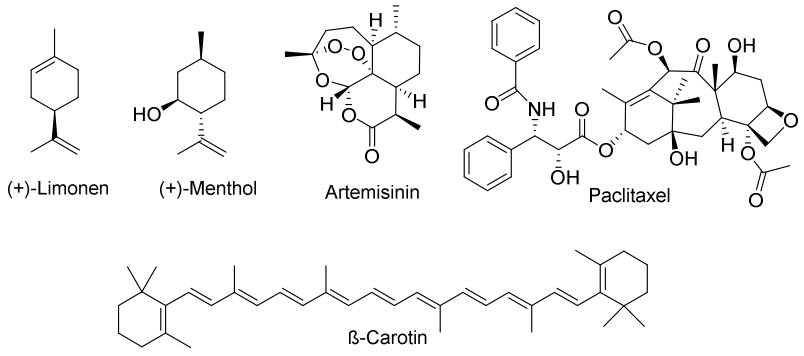


Figure 18. Examples of the terpene natural product family.

Some well-known representatives of the terpene natural product family are the plant derived monoterpenes menthol and limonene, which are industrially used as aromatic substances in the food industry.[24,132] Pharmacologically relevant examples are the anti-malaria drug artemisinin, a sesquiterpene from the plant *Artemisia annua* and the diterpene paclitaxel isolated from the pacific yew *Taxus brevifolia*.[133,134] Paclitaxel is a powerful cytoskeletal drug that targets tubulin and is applied in cancer treatment.[135] β-carotin is an example for an organic

tetraterpenoid red pigment. It is the metabolic precursor for the production of vitamin A and frequently used as a dye for food coloring.[136,137]

3.1.2 Production of Terpenoids in a Bio-Refinery

Bio-refineries can be used to fabricate a variety of products such as commodities, speciality and fine chemicals.[36,138] The production of fine chemicals is of special financial interest, since these substances generate high value even in kg scales.[139] Terpenes display high structural complexity, which often prevents economically efficient total synthesis.[8] Currently, many speciality and fine chemicals are produced by microbial fermentation including amino acids and vitamins.[32] Additionally, a wide spectrum of natural product antibiotics are produced in their natural host strains such as *Streptomyces spp*, *Bacillus spp.* and *Aspergillus spp*, which have been optimized for increased production by unspecific mutagenesis or by optimizing endogenous biosynthetic pathways.[32] However, for the production of terpenes and other natural products, native host fermentation is often impractical or results in low yields.[32] A promising approach is the heterologous reconstitution of the biosynthetic pathways in genetically well characterized microbial hosts systems such as *Escherichia coli* and *Saccharomyces cerevisiae*.[32] By heterologous reconstitution, the production rates can be significantly increased compared to the wild type producer.[140] Furthermore, the host systems *E.coli* and *S. cerevisiae* can be used to utilize diverse carbon substrates such as glycerol, glycose, lactose or xylose that are produced in bio-refineries.[141,142,143]

Whole-cell catalysis system engineered in the microbial hosts *E. coli* and *S. cerevisiae* have been demonstrated to produce terpenes in g/L scale.[9,144] For example, an engineered *E.coli* strain was modified to produce the monoterpenes limonene (400 mg/L) and perillyl alcohol (100 mg/L) from glucose.[132] The sesquiterpene artemisinic acid, a biosynthetic intermediate for the production of the anti-malaria drug artemisinin can be produced in *S. cerevisiae* with a yield of 1.6 g/L using a glucose/ethanol feed.[133] The final synthesis to artemisinin is then accomplished by 5 subsequent semisynthetic chemical steps.[133] The diterpene paclitaxel can be produced in *Tsuga chinensis* plant cell lines at rates of 137 mg/L.[145] Using genetic engineering techniques in *E.coli*, it was possible to produce the early metabolic precursor taxadiene-5-α-ol with a yield of 24 mg/L.[146] However, the complete production of paclitaxel with superior production rates compared to plant cell fermentation requires the successful reconstitution of a series of enzymes and provides a challenging research topic. [146]

3.1.3 Biosynthesis of Terpene Macrocycles

Upstream Pathway

Terpene biosynthesis has evolved in two independent pathways, which provide the universal terpene precursors isopentenyl diphosphate (IPP) and dimethylallyl pyrophosphate (DMAPP).[147] These precursors comprise a 5 carbon skeleton and are synthesized by either the mevalonate-dependent pathway (MEV) or the 1-deoxy-D-xylose-5-phosphate (DXP) dependent biosynthetic route.[147] Both pathways are depicted in figure 19 and figure 20.

The mevalonate pathway

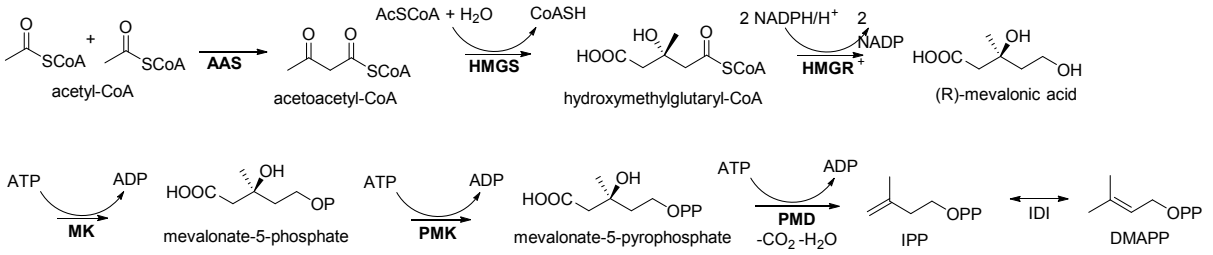


Figure 19. The mevalonate (MEV) -dependent pathway for the biosynthesis of the terpene precursors isopentenyl diphosphate (IPP) and dimethylallyl pyrophosphate (DMAPP). Enzymes: acetoacetyl-CoA thiolase (AAS), 3-hydroxy-3-methylglutaryl -CoA synthase (HMGS), 3-hydroxy-3-methylglutaryl -CoA reductase (HMGR), mevalonate kinase (MK), phosphomevalonate kinase (PKM), mevalonate-5-pyrophosphate decarboxylase (PMP).

The MEV pathway is the primary pathway for the terpene biosynthesis in eukaryotes such as plant, fungi and invertebrates.[147,148] In this pathway the primary platform metabolite is acetyl-CoA, which is derived from glucose metabolism, fatty acid degradation or catabolism of amino acids and is converted into the universal terpene precursors IPP and DMAP.[149] In the first step of the MEV pathway, two molecules of acetyl-CoA are linked by the actoacetyl-CoaA thiolase (AAS) in a Claisen condensation to yield acetoacetyl-CoA.[149] Acetyl-CoA and acetoacetyl-CoA, are then further processed to 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA) by the HMG-CoA synthase (HMGS).[149] HMG-CoA is reduced to mevalonate by the HMG-CoA reductase (HMGR) using two molecules of $NADPH/H^+$.[149] HMGR is the rate controlling enzyme in the mevalonate pathway and has been targeted by a series of cholesterol lowering drugs (statins) in human.[149] In the next step, mevalonate is phosphorylated twice, first by the mevalonate kinase (MK) and then by the phosphomevalonate kinase (PMK) to yield mevalonate-5-pyrophosphate.[149] Mevalonate-5-pyrophosphate is decarboxylated by ATP consumption to isopentenyl pyrophosphate (IPP).[149] The reaction is catalyzed by the mevalonate-5-pyrophosphate decarboxylase (PMD).[149] Eventually, IPP is isomerized by the isopentyl pyrophosphate isomerase (IDI) yielding dimethylallyl pyrophosphate (DMAPP).[149]

1-Deoxy-D-Xylose-5-Phosphate Pathway

In contrast to the MEV pathway, the 1-deoxy-D-xylose-5-phosphate (DXP) dependent pathway is found in bacteria, plant pastides and in apicomplexan protozoa such as *Plasmodium falciparum* – the malaria causing pathogen.[150,151,152] In comparison to the MEV pathway, the DXP pathway is more complex and needs 7 versus 6 steps.[9] Furthermore, it works at higher molecular and redox efficiency.[9] While in the DXP pathway 1.26 mol glucose are converted to 2.54 mol CO_2 and 1.00 mol IPP, the MEV pathway utilizes 1.50 mol glucose to yield 4.00 mol CO_2, 1.00 mol IPP and 4.00 mol NADH.[9] The DXP pathway uses the metabolic precursors acetyl-CoA and glycerolaldehyd-3-phosphate, which are provided by glycolysis or the pentose phosphate pathway.[9]

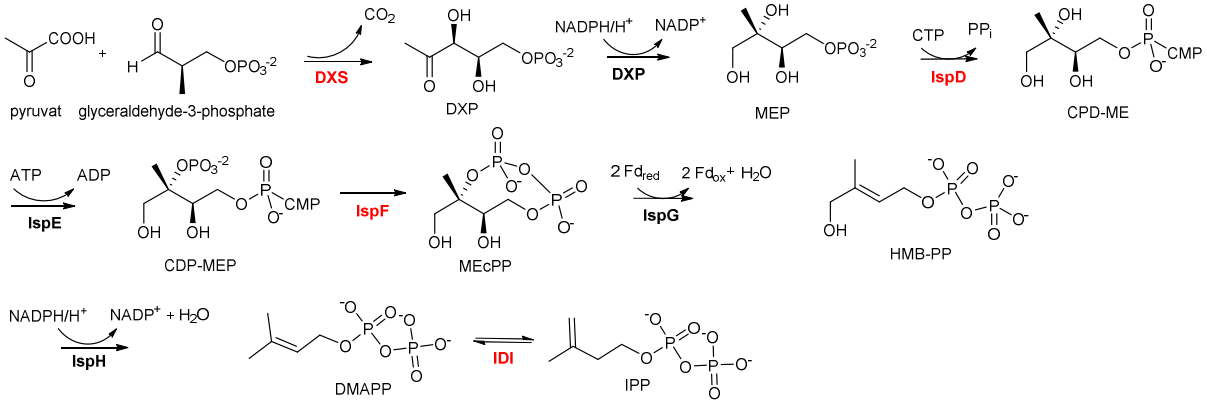


Figure 20. The 1-deoxy-D-xylose-5-phosphate (DXP) pathway for the biosynthesis of the terpene precursors isopentenyl diphosphate (IPP) and dimethylallyl pyrophosphate (DMAPP). Enzymes names: 1-Deoxy-D-xylulose 5-phosphate synthase (DXS), 1-Deoxy-D-xylulose 5-phosphate reductoisomerase (DXP), 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase (IspD), 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (IspE), 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (IspF), (E)-4-Hydroxy-3-methyl-but-2-enyl pyrophosphate synthase (IspG), the isopentyl pyrophosphate isomerase (IDI). Enzymes highlighted in red represent bottlenecks in the DXP pathway.

In the first step of the DXP pathway acetyl-CoA and glycerolaldehyd-3-phopshate are condensed to 1-Deoxy-D-xylulose 5-phosphate (DXP) by the DXP synthase (DXS).[150] DXP is then turned into 3-C-methylerythritol-4-phopshate (MEP) by the DXP reductoisomerase (DXP).[150] Next, MEP is converted by the 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase (IspD) into 4-diphosphocytidyl-2-C-methylerythritol (CDP-ME).[150] Subsequently, CDP-ME is phosphorylated by the 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (IspE) to yield 4-diphosphocytidyl-2-C-methyl-D-erythritol 2-phosphate (CDP-MEP).[150] CDP-MEP is cyclized by the 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (IspF) to 2-C-methyl-D-erythritol 2,4-cyclopyrophosphate (MEcPP), which is then reduced and simultaneously cleaved into (E)-4-Hydroxy-3-methyl-but-2-enyl pyrophosphate (HMB-PP) by the IspG synthase utilizing two electrons each from a [Fe_2S_2] cluster containing ferredoxin.[150] In

the last step HMB-PP is reduced to DMAPP, which is then isomerized to IPP by the isopentyl pyrophosphate isomerase (IDI) analogous to the mevalonate pathway.[150] Several rate limiting steps have been determined in the DXP pathway, which encompass the enzymes DXS, IspD, IspF and IdI (colored red in figure 20).[146] Overexpression of these enzymes in *E.coli* was demonstrated to result in a highly efficient (>1 g/L) terpene production system.[146]

Downstream Biosynthetic Pathway

The universal precursors IPP and DMAPP are connected into linear building block precursors by prenyltransferases.[153,154] In this reaction IPP is repeatedly added to DMAPP by cleavage of the diphosphate group.[154] Depending on the number of chemical condensation steps, the formation of geranyl pyrophosphate (GPP, C10), farnesyl- pyrophosphate (FPP, C15), geranylgeranyl pyrophosphate (GPP, C20) and farnesyl geranyl pyrophosphate (C25, FGPP) is generated.[154] These linear prenyl diphosphate building blocks are implemented as substrate in a huge variety of terpene synthases and processed through cyclisation cascades into structurally diverse terpenes with mono- or multi-cyclic carbon skeletons.[154] From a chemical perspective, the reaction cascade is initiated by the formation of a highly reactive carbocation, which propagates through the carbon skeleton connecting chemical bonds and facilitating ring closure.[130,155,156] The cyclisation reaction often involves hydride, methyl shifts, deprotonation and reprotonation. The reaction is terminated by carbocation quenching either by deprotonation or the addition of water.[130,155,156] Diterpene synthases are highly stereospecific and often catalyse the formation of several stereocenters in a single terpene molecule.[157] Most terpenoids with biological activity are often extensively modified encompassing hydroxyl, carbonyl, ketone, aldehyde, peroxide groups.[130] Functionalization of terpene carbon skeletons is most exclusively initiated by cytochrome P450 enzymes, due to their unique ability to hydroxylate non-activated carbon atoms.[144,158]

3.1.4 Terpene Synthases

Terpene synthases are distinguished by the mechanism used for the carbocation formation.[159] Class I terpene synthases generate an allylic carbocation by cleavage of the pyrophosphate group.[159] This class contains a conserved DDXXD and NSE/DTE motif that binds the pyrophosphate group by coordination of a trinuclear Mg^{2+} ion cluster.[159] In contrast, class II terpene synthases use an acidic amino acid to protonate the terminal isoprene unit, leaving the pyrophosphate unaffected.[159] Protonation is facilitated by a conserved DXDD motif. These

mechanistic differences are reflected in the protein folding (Figure 21).[159] Bacterial class I terpene synthases are structurally related to isoprenyl diphosphate synthases such as the human farnesyl diphosphate synthase and display a single α-helical domain (Figure 21a).[159] Bacterial class II terpene synthases display a multi-domain architecture containing the α-helical domains (βγ), which were probably derived from an ancient duplication of two $(\alpha/\alpha)_6$ barrels (Figure 21b).[159] The β-domain is structurally similar to a β-subunit found in plant and animal prenyl transferases, where it acts as binding site for FPP or GGPP (Figure 21d).[159] The N-terminus of the γ-domain in the βγ class II terpene synthase folds back to the β-domain and is illustrated in blue (Figure 21b).[159]

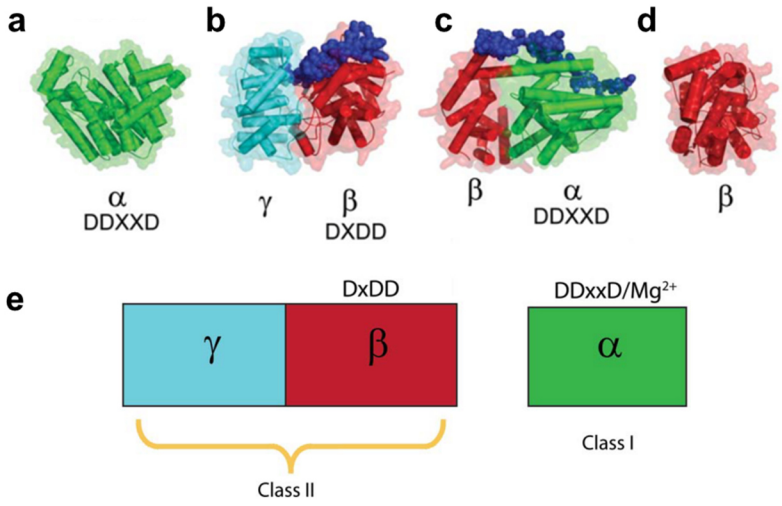


Figure 21.[159] The different terpene synthase folds are illustrated. (**a**) α domain (green) of the human farnesyl pyrophosphate synthase harboring the catalytic DDXXD domain. (**b**) βγ domains in the bacterial triterpene squalene hopene synthase containing the DXDD motif in the β domain (red). The N-terminus is illustrated in blue and folds back from the γ to the β domain. (**c**) αβ domains of the class I plant bornyl diphosphate synthase. The α domain houses the DDXXD motif. The N-terminus folds back from the β domain and is shown in blue .(**d**) A protein geranylgeranyl transferase displaying the β domain as $(a/a)_6$ barrel. (**e**) Cartoon diagram showing all three domains (α, β,γ) with their corresponding activities including key catalytic residues.

Plant terpene synthases are structurally more complicated.[159,160] Class I mono and sesquiterpene synthases from plants contain an additional β-helical domain that is fused to the N-terminal end of the α-domain (Figure 21c).[159] By contrast, plant class I/II diterpene synthases are constructed from a αβγ-tridomain structure, which have probably evolved from ancient fusion of bacterial class I and class II synthases.[159] These bifunctional terpene synthases can either contain a single functional α-domain, which facilitates class I cyclisation of diterpenes, or cyclisation may involve all three domains.[159] In the latter case, cyclisation starts by a class II cyclisation reaction in the βγ-domains resulting in a stable bio-catalytic intermediate e.g. copalyl diphosphate and ent-copalyl diphosphate.[159] This intermediate is subsequently

transferred to the α domain, where cleavage of the pyrophosphate group initiates cyclisation to the complete diterpene structures e.g. ent-kaurene and pimaradiene (Figure 22).[159]

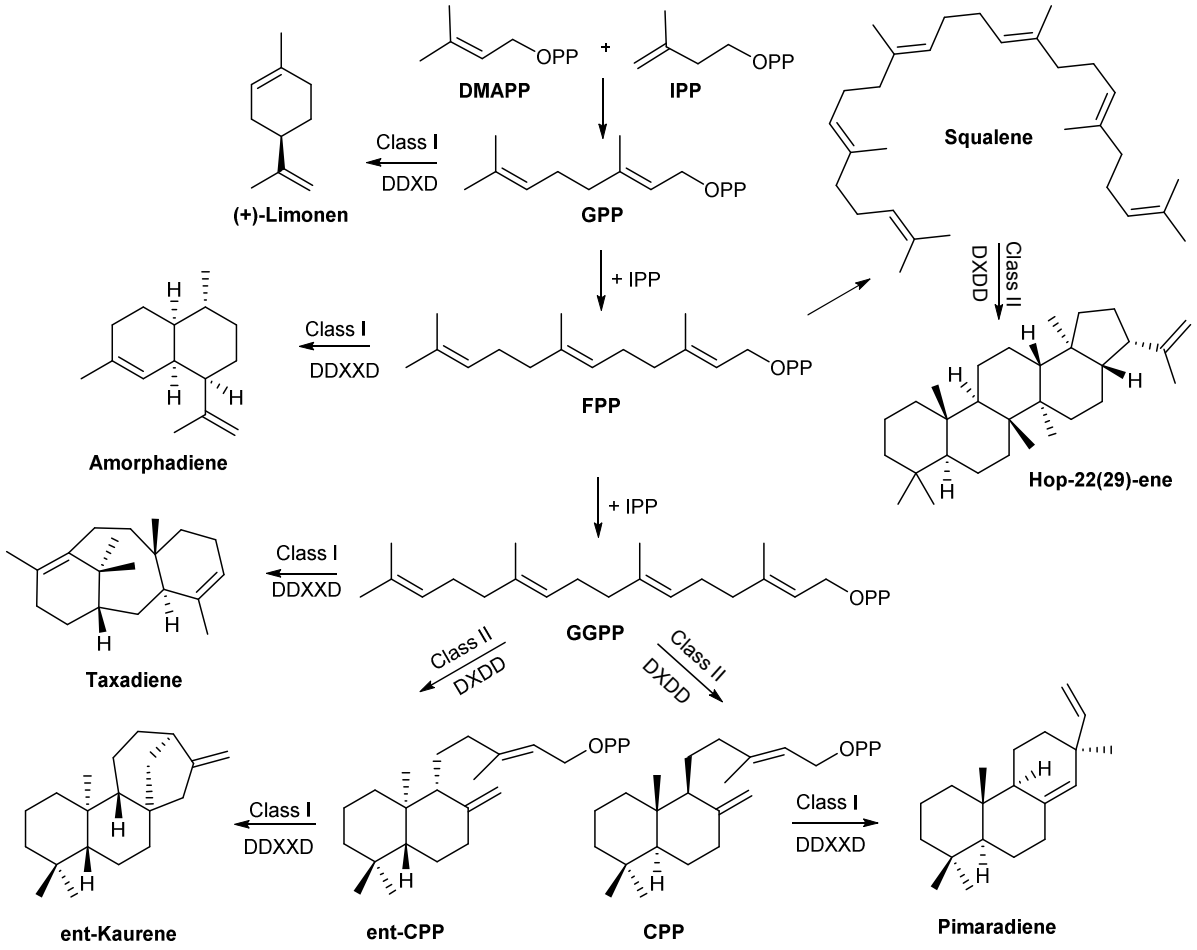


Figure 22. Overview of terpene biosynthesis. Figure adapted from [159]

The largest portion of terpene synthases has been isolated from fungi and plants. Since only few eukaryotic genomes have been sequenced, isolation by cDNA libraries is the preferred method. Despite the high structural similarities found in class I and class II terpene synthases, bacterial terpene synthases show surprisingly low primary sequence similarity.[161,162] The lack of sequence conservation challenges a sequence based discovery.[161,162] As a result, many bacterial terpene synthases were identified by clustering with more easily identifiable terpene biosynthetic genes.[162] Particularly, the identification of GGDP synthases in an operon has revealed the existence ditperene synthases.[162]

3.1.5 The Cyclooctatin Gene Cluster

The diterpene cyclooctatin belongs to the fussicoccane natural product family, which comprise diterpenes featuring a 5-8-5 ring motif that are found in bacteria, fungi and plants.[163] Depending on the functionalization, fussicoccane display a broad biological activity encompassing ATPase activation, bacteriostatic, fungicidal, and tumour static effects.[163] Cyclooctatin is a lysophospholipase inhibitor with potential anti-inflammatory activity.[13] Lysophospholipases catalyse the hydrolysis of fatty acid ester bonds in lysophospholipids and are upregulated in eosinophilic leukocytes that occur in allergic reactions and inflammatory diseases.[164] Recently, the cyclooctatin gene cluster from the soil bacterium *Streptomyces melanosporofaciens* MI614-43F2 was characterized comprising an operon with four open reading frames, the GGPP synthase CotB1, the diterpene synthase CotB2 and two cytochrome P450 CotB3/ CotB4.[10] The biocatalysis of cyclooctatin is accomplished by the CotB2 mediated cyclisation of the GGPP precursor to cyclooctat-9-en-7-ol, which is further hydroxylated by the P450 cytochrome CotB3 to cyclooctat-9-en-5,7-diol and by CotB4 to cyclooctatin (Figure 23).[10]

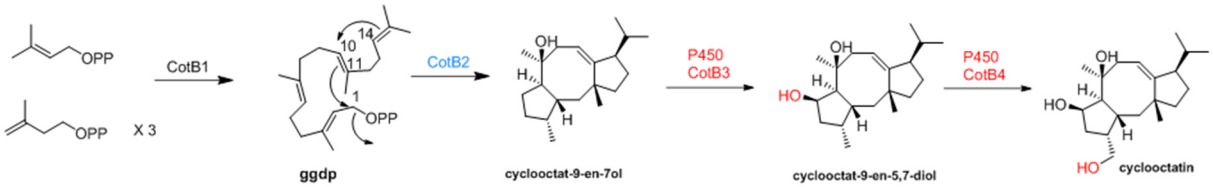


Figure 23. The cyclooctatin biosynthesis in *Streptomyces melanosporofaciens* MI614-43F2. Cyclooctat-9-en-7-ol is the cyclisation product of the diterpene synthase CotB2. Subsequent functionalization by the P450 monoxygenases CotB3 and CotB4 results in the production of cyclooctatin.

3.2 Scope of This Work

Diterpenes belong to the terpene family and are characterized by a macrocycle consisting of 20 carbon atoms.[7,8] Diterpenes may exhibit diverse biological effects such as antimicrobial, antifungal, anti-inflammatory and cytostatic activity.[9] The chemical synthesis of diterpene macrocycles is challenging and is often characterized by low yields, the formation of unfavorable isomers and toxic waste streams.[8] In contrast, the biochemical production of diterpenes macrocycles by diterpene synthases is done in a highly selective single step transformation using geranylgernyldiphosphate (GGPP) as substrate.[7,8]

The biotechnological exploitation of this route requires the identification and successful reconstitution of diterpene synthases in a suitable production host.[24] To-date only a limited number of diterpene synthases has been characterized, leaving the majority of the diterpene macrocycle diversity only accessible by classical natural product synthesis.[165] In contrast to the discovery of diterpene synthases, protocols have been developed for the rapid identification of mono and sesquiterpenes synthases. The identification is based on the volatility of mono and sesquiterpenes, which can readily be detected by head space GC-MS analysis thus avoiding time consuming extraction and purification steps.[166] In this approach, gene libraries harboring potential terpene synthases are assayed by high-through-put screenings.[166] By contrast, the identification of non-volatile diterpenes and the corresponding diterpene synthases is time consuming and involves many preparation and purification steps.

The first part of this study focuses on the engineering of the bacterial cyclooctat-9-en-7-ol diterpene synthase (CotB2) for the production of non-native diterpenes that are lacking awareness of a corresponding synthase. The combination of *in silico* modelling and crystallographic studies is used for the targeted engineering of CotB2. The enzyme engineering approach applied in this study resembles an alternative strategy to access synthases that produce novel diterpenes, compared to classical genomical and proteomical approaches.

Bacterial diterpene synthases represent a valuable engineering target by displaying a single α-helical domain structurally similar to mono and sequiterpene synthases.[162] Furthermore, these enzymes can be produced effectively in standard bacterial expression systems such as *E.coli*.[162] Sesquiteperpene synthases have been extensively engineered to produce novel carbon skeletons and to elucidate structure-function relationships. Examples are the pentalenene and

aristolochene synthase.[167,168] These engineering efforts led to a variety of different macrocycles compared to the wild type cyclisation product.[167,168] By contrast, only few reports describe the engineering of diterpene synthase. In this regard, mutations have been reported to alter the cyclisation cascasde in plant class I diterpene synthases such as the pimaradiene and ent-kaurene synthase.[169,170,171,172,173] These enzymes utilize the two ring intermediates copalyl (CPP) and ent-copalyl pyrophosphate (ent-CPP) as substrates. Accordingly, mutations in the diterpene synthase only gave wild type related diterpenes structures.[169,170,171,172,173] A comprehensive alteration of the carbon skeleton, which has been demonstrated in sequiterpenes synthases has not been reported for diterperene synthases and is demonstrated in this study for the first time.

The second part of this study investigates the catalytic mechanism of the native and mutant cyclooctat-9-en-7-ol diterpene (CotB2) synthase. Therefore, CotB2 variants that produced new diterpene are analyzed by *in silico* modelling to elucidate the structure-function relationship for the production of new diterpenes.

3.3 Result and Discussion

3.3.1 Modelling and Crystal Structure of the Bacterial Diterpene Synthase CotB2

In silico Structure Prediction

The *in silico* structure prediction of CotB2 was done by Patrick Schrepfer and is part of the paper: "Targeted Engineering of Cyclooctat-9-en-7-ol Synthase: A Stereospecific Access to Two New Non-natural Fusicoccane-Type Diterpenes" Görner, C.; Häuslein, I.; Schrepfer, P.; Eisenreich, W.; Brück, T. *ChemCatChem* **2013**, *5*, 3289–3298.

At the beginning of the study a crystal structure of CotB2 and any other bacterial diterpene synthase was unavailable. To choose amino acids for mutagenesis, an *in silico* structural model of CotB2 was created (Figure 24).

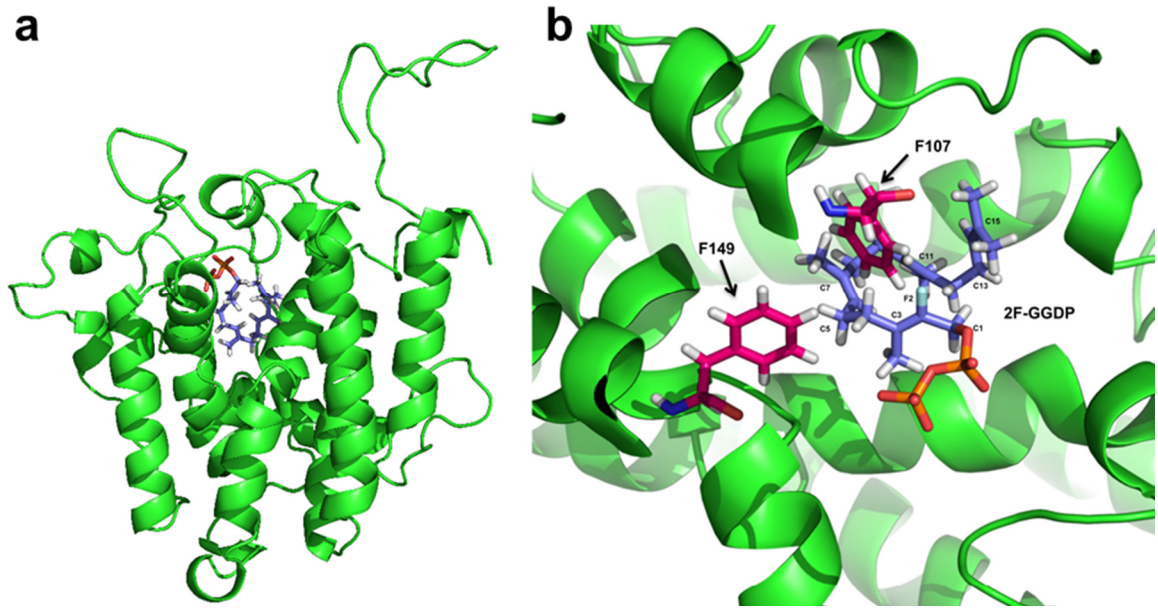


Figure 24. (**a**) Comprehensive view of the *in silico* created model of CotB2 harboring the isoprenyl substrate buried within the helical structures. (**b**) Detailed view of the catalytic center of CotB2. The substrate 2F-GGDP (2-fluoro-geranylgeranyldiphosphate) from the crystal structure of the taxadiene synthase is in close proximity to the catalytic residues F107 and F149.

The amino acid sequence was used as template for a sequence homology detection and secondary structure prediction by Hidden Markov Model comparison using the HHpred server.[174] CotB2 displayed the highest similarity to the fungal aristolochene synthase, a sesquiterpene synthase from *Aspergillus terreus* (pdb ID: 3BNY) and resulted in an E-value of 2.6×10^{-12}. Next, the program modeller[174] was used for structure modelling based on the aristolochene synthase as template with a respective E-value of 8×10^{-5}. To identify essential amino acids in close proximity to the GGPP substrate the isoprenyl precursor was adapted from

the structural data of the taxadiene synthase from the *Pacific yew*.[175] Therefore, tertiary structural alignment of the CotB2 model with the α-domain of the taxadien synthase was carried out by the MUSTANG algorithm of the Yasara bioinformatics suite. The model of the bacterial CotB2 showed significant similarity with the plant derived α-domain of the taxadiene synthase. The alignment resulted in a r.m.s.d (root mean square deviation) value of 0.291 Å over 162 aligned amino acid residues with an overall 8.64% primary sequence identity. Due to the evolutionary relationship of prokaryotic synthase with the α-domain of plant diterpene synthases, a high similarity in the barrel fold and positioning of the helical strands is not surprising. Analysis of the CotB2 model allowed the identification of amino acids F107 and F149 with potential catalytic activity (Figure 24b).

Crystallographic Structure Elucidation of CotB2[164]

The crystal structure of CotB2 was prepared and solved by Ronja Janke and has been published along with the characterization of CotB2. See: Janke, R.; Görner, C.; Hirte, M.; Brück, T.; Loll, B. *Acta Crystallogr. D. Biol. Crystallogr.* **2014**, *70*, 1528–1537.

In the course of this study, it was possible to obtain more detailed structure information from the crystal structure of CotB2 by a collaboration with Ronja Janke (FU Berlin) [164] CotB2 crystalized at 1.64 Å resolution in the open and inactive configuration missing a substrate or substrate analog (Figure 25).[164]

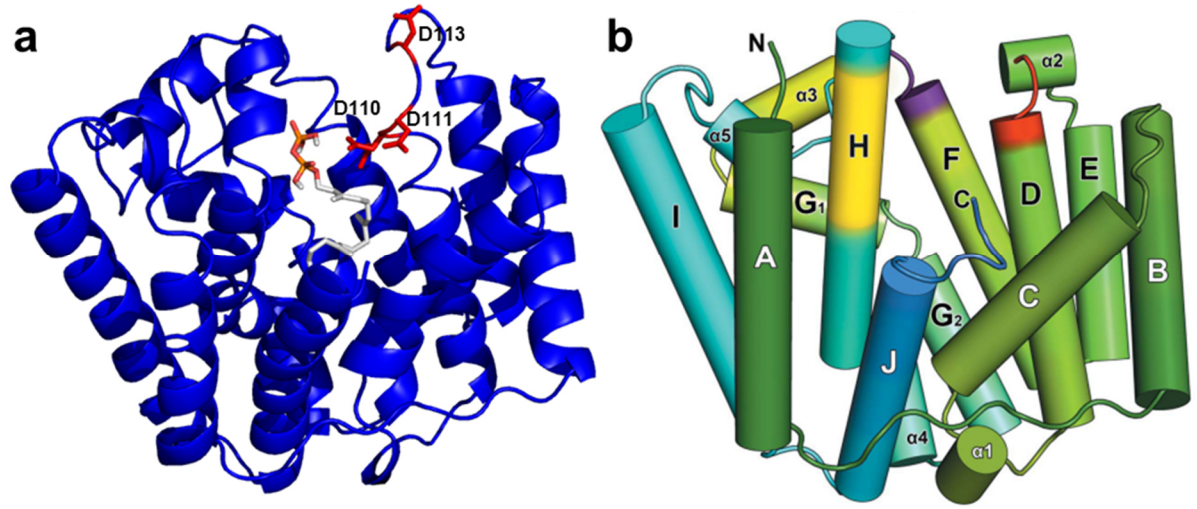


Figure 25. (**a**) Crystal structure of the CotB2 monomer illustrated in blue harboring the *in silico* docked GGDP substrate (gray). The pyrophosphate group is colored in orange. The side chains of Asp 110, Aps 111 and Asp 113 are highlighted in red. Asp 113 is distinct from Asp 110 and Asp 111 and does not participate in binding of the GGDP substrate. (**b**) CotB2 monomer displayed in a color gradient from green (N-terminus) to blue (C-terminus). The α-helices are drawn as cylinders. The aspartate-rich motif is highlighted in red, the NSE/DTE motif is marked in yellow. The double conformation at the C-terminal end of the α-helix F is indicated in purple.

The X-ray structure was solved as a protein dimer and was missing 15 N-terminal and 10 C-terminal amino acid residues. The monomeric CotB2 consists of 10 core α-helices (A-J) (Figure

25b) that are connected by short loop segments and 5 short α-helices (α1- α5). The active site is surrounded by the core helices and buried at the center of the enzyme. Interestingly, the structurally closest neighbours were found in the plant derived synthase 4S-limonene (α-domain) as well as the fungal sesquiterpene synthases trichodiene, aristolochene and the bacterial sesquiterpene synthase epi-isoziziaene.[176,177,178,179] All structure showed a r.m.s.d. of 3.3–3.6 Å in comparison with CotB2. Recently, the crystal structure of the sesquiterpene selinadiene synthase was published, which shows the so far closed structural homology displaying a r.m.s.d. of 2.171 Å.[180] Interestingly, the amino acid sequence of CotB2 contains an altered ^{110}DDXD^{113} class I motif instead of the classical DDXXD motif. The DDXD aspartate rich motif is located at the C-terminal end of the α-helix D and displays some unusual structural elements. In comparison to all other class I terpene synthases, the α-helix D is shorter. In this regard, a proline amino acid after the DDXD motif introduces a break in the helix that causes Asp 113 to point away from the active center (Figure 25a)

The structure function relationship of the sesquiterpene producing pentalenene synthase has been reported.[167] In this study, mutation of the aspartate residues in the DDXXD motif to glutamate resulted in rearrangement of the FPP substrate leading to the formation of various new products.[167] This promising approach was applied to CotB2. Unfortunately, functional analysis of CotB2 D110E and CotB2 D111E did not lead to the formation of new products. Instead, enzyme inactivation was observed, which implies an important contribution of these residues for the Mg^{2+} facilitated binding of the pyrophosphate group. In agreement with the odd position of Asp 113, which was expected not to coordinate Mg^{2+}(Figure 25a), the CotB2 D113E indeed displayed wild type activity.

The NSE/DTE motif in CotB2 is located in the amino acid sequence ^{218}IVNDFYSYDRE^{228} in the α-helix H. In particular, N220 from the NSE/DTE motif is pointing into the active site. Both the DDXXD and NSE/DTE motifs are essential for coordinating the Mg^{2+} ions as well as binding and cleaving of the pyrophosphate group.[177] Interestingly, at the C-terminal end of helix F, the protein backbone of the sequence (^{155}MFRD^{158}) adopts a double conformation. Furthermore, several amino acid side chains in this segment are displayed as rotamers. As a consequence, the structural flexibility of helix F, indicated by the crystal structure, may be essential for a structural rearrangements upon substrate binding. According to the crystal structure, the hydrophobic cavity that creates the active site is decorated by the residues Phe77, Val80, Phe107, Trp109, Phe149, Val150, Ile181, Phe185, Met189, Trp186, Leu281 and Trp288.

3.3.2 Mutagenesis Study of CotB2

The mutagenesis studies presented here were executed under my guidance in the Master's theses "Biotechnologische Produktion und Modifikation neuer Diterpene" 2014 from Max Hirte and „Mechanistische Studien an der Diterpenzyklase CotB2 durch Produktion und Charakterisierung neuer Diterpene" 2013 from Ina Häuslein as well as the Bachelor's thesis "Produktion, Isolation und Charakterisierung von neuen Diterpenen" 2013 from Anna Maike Beckenkamp. The data was adapted for this chapter.

In the beginning of this study only *in silico* structure derived informations were available, which were complemented by the crystal structure in the subsequent process of this study. To screen for alternative cyclisation products, a target oriented saturation mutagenesis was chosen. Hence, PCR based gene libraries were created each consisting of a randomized codon. Due to the lack of a high through-put screening system to analyse diterpene cyclisation products, the gene libraries were kept intentionally small to cover a large part of the library. The gene libraries were initially analyzed by an *in vitro* assay, which was later replaced by an *in vivo* assay. Detection of the terpene products from mutant CotB2 was carried out by GC-MS analysis.

For the *in vitro* assay, the gene libraries were harbored by the pET24a plasmid carrying a codon optimized *cotb2* gene. After transformation in *E.coli*., up to 64 clones containing CotB2 mutants were cultivated in small scale flasks for protein expression. Subsequently, the soluble *E.coli* protein fraction containing the Cotb2 variants was incubated with the GGPP precursor and the reaction products were eventually analyzed by GC-MS. The in *vivo assay* consists of an engineered *E.coli* strain overexpressing bottleneck enzymes of the DXP pathway as shown by Schrepfer et. al.[181] Therefore, the DXP upstream enzymes 1-Deoxy-D-xylulose 5-phosphate synthase (DXS), 1-Deoxy-D-xylulose 5-phosphate reductoisomerase (DXP), 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase (IspD), 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (IspF) and the isopentyl pyrophosphate isomerase (IDI) were overexpressed in *E.coli* Bl21 (DE3) using the strong T7 promotor. Furthermore, the downstream pathway enzymes geranylgeranyl diphosphate synthase (CrTE) from the bacteria *Pantoea agglomerans* and CotB2 were overexpressed. In total the engineered *E.coli* Bl21(DE3) strain was carrying three plasmids consisting of pColaDuet-1(*dxs*/*dxp*), pCDFDuet-1(*ispD*/*ispF* and *idi*) and plasmid pETDuet-1 (*crtE*/*cotB2*).

For *in vivo* screening of the gene library, 30-40 *E.coli* strains expressing CotB2 variants were cultivated. The culture medium was extracted by organic solvents and was analyzed by GC-MS. In comparison to the *in vitro* assay, the *in vivo* assay was slightly less sensitive but omitted the

use of the expensive diterpene precursor GGDP. Gene libraries were created for the amino acids Y77, F107, F149, W186, W288 and N285, which are all buried deeply in the active site and thought to be important for the cyclisation reaction according to the *in silico* model and crystal structure (Figure 26). The gene libraries for the amino acids F107, F149, and W186 were screened by the *in vitro* assay. Libraries for the amino acids Y77 and W288 were screened by the *in vivo* assay. The amino acid N285 was screened using *in vitro* and *in vivo* assays, respectively.

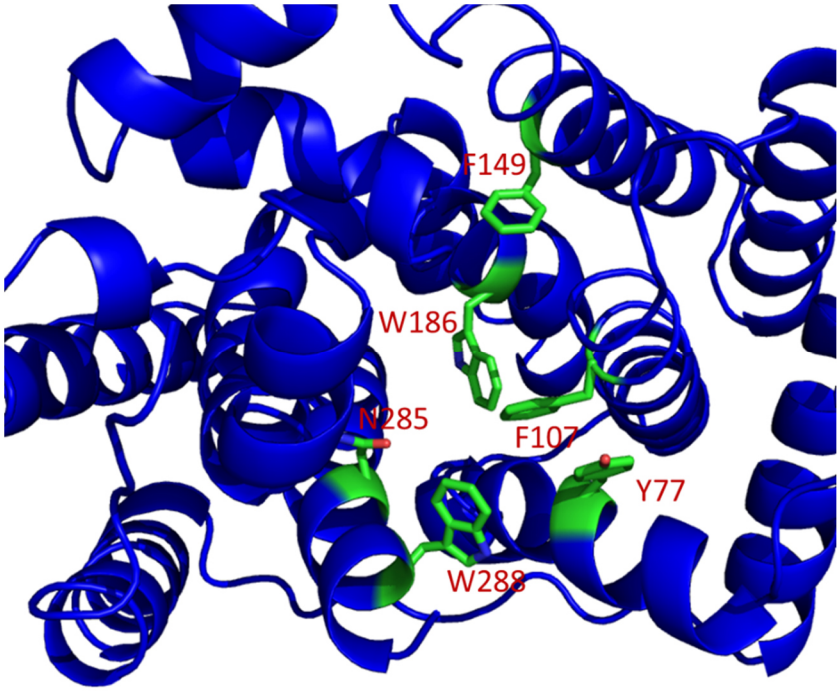


Figure 26. Catalytic center of CotB2 illustrated in blue. Mutated amino acid residues are highlighted in green. The CotB2 structure is derived from the open complex crystal structure.

3.3.3 Mutants that Generate New Macrocycles

The spectroscopic data presented here was recorded under my guidance in the Master's theses "Biotechnologische Produktion und Modifikation neuer Diterpene" 2014 from Max Hirte and „Mechanistische Studien an der Diterpenzyklase CotB2 durch Produktion und Charakterisierung neuer Diterpene" 2013 from Ina Häuslein as well as the Bachelor's thesis "Produktion, Isolation und Charakterisierung von neuen Diterpenen" 2013 from Anna Maike Beckenkamp. The data was adapted for this chapter.

After screening of the libraries, it was found that only few mutants produce novel diterpene macrocycles in amounts which were sufficient for further structural elucidation studies.

Mutations of F107 produced three altered products, indicating the importance of this amino acid residue in the cyclisation process. The Mutant F107Y resulted in a mixture of two new macrocycles, while F107G/A gave a single novel product (Figure 27). Single cyclisation products were also detected for the mutant F149/G/V/L/H and W288G, respectively (Figure 27). However, in comparison to the aforementioned mutants, CotB2 W288G showed the least product promiscuity, which is concluded by the appearance of small amounts of side products including the wild type product on the GC-MS chromatogram (Figure 27).

The MS spectrometry data of the mutant diterpene macrocycles was compared with the NIST08MS spectral database library containing a special natural product extension[182], but was inconclusive. Only the mutant F107A indicated the production of a monocyclic cembrene type ring system. As shown in figure 27, all mutants showed significantly different retention times in gas chromatography and displayed a completely different MS spectrum compared to the wild type product cyclooctat-9-en-7-ol. This finding indicated major structural rearrangements in the novel created products. The molecular mass of 290 Da for the F149 indicated the formation of an alcohol analog to the wild type product. In contrast, products of the F107 and W288 mutants displayed the molecular mass of 272 Da and concluded the absence of an alcohol function and the presence of an additional olefin bond.

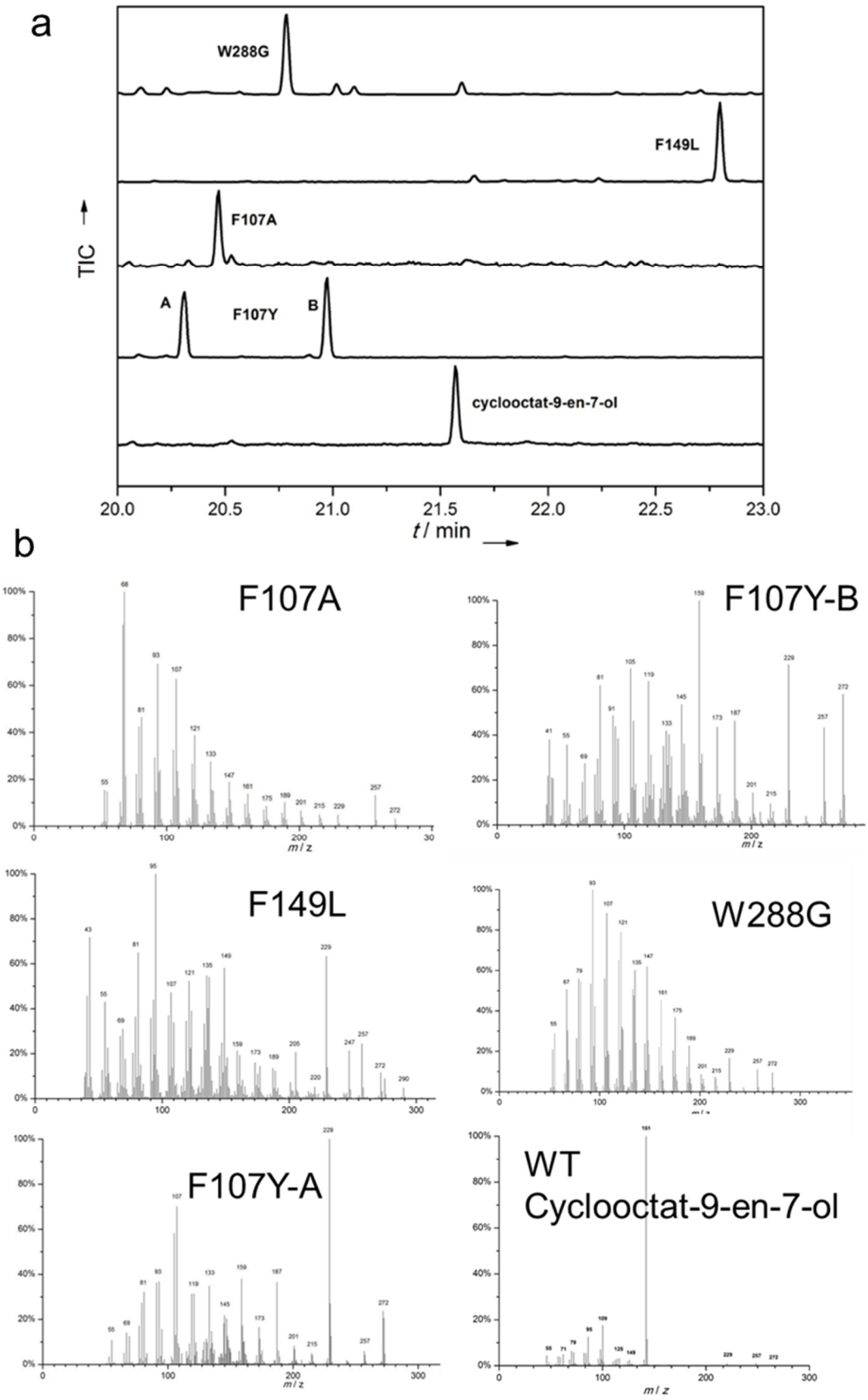


Figure 27. (**a**) GC-MS analysis of CotB2 mutants and wild type (cyclooctat-9-en-7-ol) with the corresponding MS spectra (**b**).

3.3.4 Structure Elucidation of Novel Diterpene Products

The spectroscopic data presented here was recorded under my guidance in the Master's theses "Biotechnologische Produktion und Modifikation neuer Diterpene" 2014 from Max Hirte and „Mechanistische Studien an der Diterpenzyklase CotB2 durch Produktion und Charakterisierung neuer Diterpene" 2013 from Ina Häuslein as well as the Bachelor's thesis "Produktion, Isolation und Charakterisierung von neuen Diterpenen" 2013 from Anna Maike Beckenkamp. The data has been re-analyzed, re-evaluated and re-interpreted for this chapter.

CotB2 F107A

For the structural elucidation of the novel diterpene macrocycle, larger quantities were produced by using the aforementioned engineered *E.coli* strain applied in the *in vivo* screening assay. Therefore, the cultivation volume was upscaled to several litres. Diterpene products were extracted from the cell biomass and cultivation medium, purified and characterized by detailed NMR spectroscopy (Appendix Figure 29, 30)

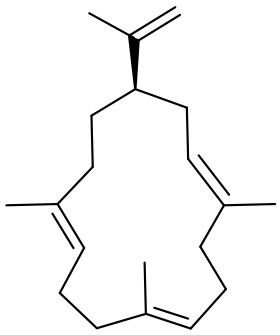


Figure 28. Chemical structure of R-cembrene A

When the biotechnological production of the CotB2 F107A product was attempted, which was previously characterized by the *in vitro* assay, it was observed that unlike any other diterpene macrocycle only very small quantities could be generated. According to literature, cembrenes may show weak stability even at low temperatures and degrade within hours of their isolation.[183] Since the biotechnological production takes up to 5 days, increasing the production yields is challenging.[183] This problem was eventually solved by adding C-18 reverse phase beads to the cell culture. The hydrophobic surface of the beads presumably adsorbs and stabilizes terpene macrocycles. However, in order to prevent saturation of the beads by the growth media components, the media composition was changed from rich to minimal medium. After cultivation the terpenes were recovered from the beads and characterized. This method gave access to about 1 mg purified F107A product from 3 times 2 liter shake flask cultures

grown for 5 days at 28°C. ^{13}C NMR spectroscopy showed the signals of 20 carbon atoms as follows: (126 MHz, $CDCl_3$) δ=149.44, 134.98, 134.05, 133.59, 126.06, 124.17, 121.94, 110.29, 46.10, 39.56, 39.09, 34.08, 32.55, 28.27, 25.03, 23.89, 19.45, 18.18, 15.68, 15.45. ^{13}C, DEPT 90/135 NMR spectroscopy indicated the presence of four olefinic bonds harboring five olefinic protons including two protons from a terminal alkene. One carbon atom showed a tertiary carbon signal, which was a hint for a monocyclic structure. Comparison of the observed ^{13}C NMR signals with reported cembrene like diterpenes was very consistent for cembrene A (Figure 28).[183,184] This finding was further confirmed by comparison of the associated ^{1}H NMR spectrum. The proton chemical shifts revealed characteristic olefinic proton signals at (500 MHz, $CDCl_3$) δ=4.74 (s, 1H) and 4.68 (s, 1H) as singlets as well as triplet signals at δ=5.22 [t, 1H, ^{3}J(H,H)=7.6 Hz, 1H], 5.00 [t, ^{3}J(H,H) = 6.2 Hz, 1H] and 5.08 [t, ^{3}J(H,H) = 6.2 Hz, 1H], which are in accordance with ^{1}H NMR signals from cembrene A.[183] As a monomeric diterpene macrocycle, cembrene A displays only one stereocenter. Consequently, the resulting enantiomers are indistinguishable by NMR spectroscopy. The absolute configuration was determined by CD spectroscopy (Appendix Figure 76). Cembrene A displayed a negative circular dichroism and was ascertained to be (-)-R-cembrene A. Interestingly, all cembrene A natural product isolates from plant, corals and bacteria to date represent the R enantiomer.[183,185,186,187]

CotB2 W288G

For the structure elucidation of the CotB2 W288G product, no prior structural information was provided by database comparison of the MS fragmentation pattern. 8 mg/L of the CotB2 W288G product were obtained from cultures supplemented with C-18 reversed phase beads. ^{13}C, DEPT 90/135 NMR spectroscopy indicated the presence of three olefinic bonds displaying four olefinic protons from which two protons were attached to a terminal olefinic bond (Table 3). Furthermore, the presence of two tertiary carbon signals was a hint for a two membered ring structure. ^{1}H NMR spectroscopy showed the presence of four singlet methyl groups and two striking broad olefinic singlet protons, which were assigned to the terminal olefinic double bond. The connectivity of the carbon skeleton was deduced by COSY, HSQC and HMBC NMR spectroscopy (Appendix Figure 43-48) and is displayed in figure 29.

Table 3. NMR data of 3,7,18-dolabellatriene.

#	δ_H (ppm), *J* (Hz)	δ_C (ppm)		#	δ_H (ppm), *J* (Hz)	δ_C (ppm)	
1	-	46.67	C	**11**	1.75 (m, 1H)	41.84	CH
2	α 2.23 (m, 1H) β 1.71 (m, 1H)	43.28	CH_2	**12**	2.63 (dt, *J* = 10.3; 6.7, 1H)	50.93	CH
3	5.14 (dd, *J* = 11.4; 4.7, 1 H)	125.77	CH	**13**	a 1.65 (m, 1H) b 1.58 (m, 1H)	28.43	CH_2
4	-	134.68	C	**14**	a 1.53 (m, 1H) b 1.43 (m, 1H)	41.97	CH_2
5	a 2.22 (m, 1H) b 2.07 (m, 1H)	39.78	CH_2	**15**	1.08 (s, 1H)	24.24	CH_3
6	a 2.29 (m, 1H) b 2.06 (m, 1H)	24.49	CH_2	**16**	1.53 (s, 1H)	15.77	CH_3
7	4.86 (m, 1H)	127.37	CH	**17**	1.53 (s, 1H)	16.78	CH_3
8	-	134.06	C	**18**	-	147.04	C
9	a 2.14 (m, 1H) b 1.89 (m, 1H)	37.42	CH_2	**19**	1.72 (s, 1H)	23.52	CH_3
10	a 1.34 (m, 1H) b 1.24 (m, 1H)	24.31	CH_2	**20**	α 4.83 (brs, 1H) β 4.65 (brs, 1H)	112.32	CH_2

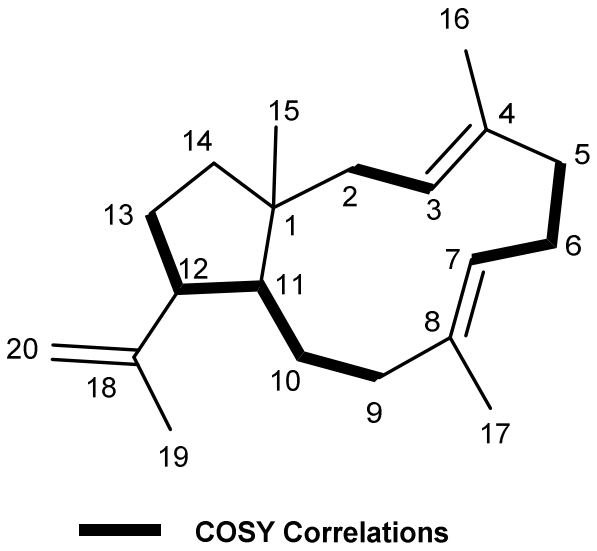


Figure 29. Structure of 3,7,18-dolabellatriene. Key correlations in the COSY NMR spectrum are indicated.

Key COSY correlations used for the structure determination are illustrated. The CotB2 W288G diterpene displayed a dolabellane type macrocycle (3,7,18-dolabellatriene) harboring a 5 and 11 membered ring system (Figure 29). The five membered ring structure was elucidated by the COSY spectrum illustrating the spin coupling system between H-11/H-12, H-12/H_2-13 in combination with the HMBC correlation between C-1/H-11, C-14/H-11, C-1/H_2-13 and C-

14/H_2-13. The 11 membered ring structure was assembled by the two spin systems H_2-2/H-3; H_2-5/H_2-6, H_2-6/H-3 and H_2-9/H_2-10, H_2-10/H-11 determined in the COSY NMR spectrum, which were connected by the HMBC correlations C-3/H_2-5, C-4/H_2-5, C-3/H_3-16, C-4/H_3-16, C-7/H_2-9, C-7/H_3-17, C-8/H_2-9, C-8/H_3-17. The 5 and 11 membered rings were fused together according to the HMBC correlations C-1/H_2-2, C-1/H_2-10 and C-11/H_2-2. The methyl group H_3-15 was assigned to C-1 according to the HMBC correlations C-1/H_3-15, C-11/H_3-15 and C-14/H_3-15. The broad olefinic singlet protons seen in the ^{1}H NMR were assigned to a single carbon atom by the HQSC spectrum. HMBC correlations between C-20/H_3-19 and C-18/H_2-20 and C18/H_3-19 indicated the presence of an isopropylene group, which was connected to C-12 by the HMBC correlation C-18/H_2-12.

Dolabellanes are a large class of natural products and are frequently found in marine algae, sponges and terrestrial plants such as liverworts.[188,189,190,11,191,192] Based on the extensive conducted research on dolabellanes, the new product was searched in the structure database Reaxys.[193] Interestingly, a hit appeared that was referenced to a previously reported natural product isolate from the brown algae *Dilophus spiralis*.[11] Comparison of the ^{1}H and ^{13}C chemical shifts with the natural product isolate were in very good agreement.

3,7,18-dolabellatriene contains three stereocenters, which connect both ring systems as well as the isopropylene function with the 5 membered ring. Furthermore, *E*-*Z* configurations are possible at the double bonds between C-3/-C-4 and C-7/C-8. The configurations of the double bonds were readily determined as 3*E*, 7*E*, based on the ^{13}C chemical shifts of the connected methyl groups H_3-16 and H_3-17, which was consistent with the natural product isolate. Chemical shifts below 20 ppm have been exclusively assigned to double bonds with *E*-configuration in dolabellane diterpenes, while methyl groups associated to double bonds with *Z*-configuration resonated at higher frequencies.[11,191,194,195]

To elucidate if the relative configuration of the stereocenters is consistent with the natural product isolate, the W288G diterpene product was analyzed by NOSEY NMR spectroscopy. Key NOE enhancements were observed between the protons H-2α/H-11, H-2β/H_3-15, H-11/H-12, H-12/H-20α, H_3-19/H-20β and H_2-13/H-20β. The NOE correlations were identical to the reported structure and indicate a *trans*-fusion of both ring systems. Furthermore, these data suggested that the methyl group H_3-15 and the isopropylene group are *cis*-oriented. The strong NOE correlations were verified for plausibility by an *in silico* created structural model. The model was created using the molecular mechanics (MM2) force field from the software

Chemdraw V12. All key NOE correlations were in agreement with the structural model and are illustrated in figure 30.

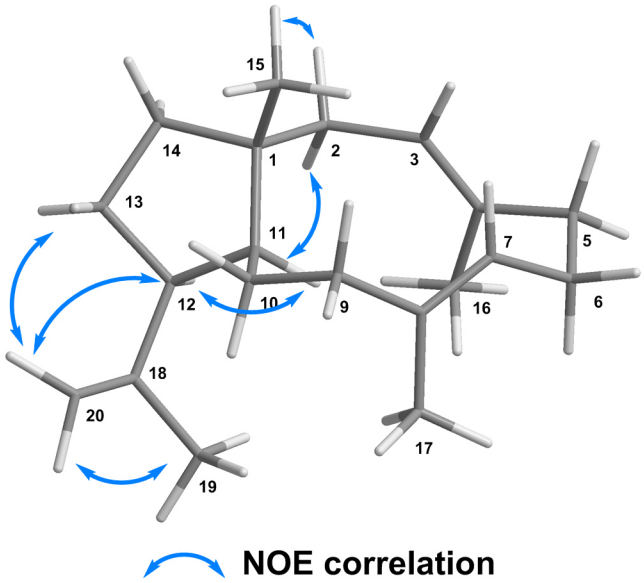


Figure 30. Molecular model (MM2) of (1R, 3E, 7E, 11S, 12S)-3,7,18-dolabellatriene. Key NOESY correlations are highlighted.

The absolute configuration was determined by CD spectroscopy (positive circular dichroism) confirming the configuration of (1R,3E,7E,11S,12S)-3,7,18-dolabellatriene as identical to the reported brown alga isolate (Appendix Figure 76). Interestingly, this natural product showed antibiotic activity against multidrug-resistant *Staphylococcus aureus* strains [MIC: 16- 128 µg/mL] without the need for further fuctionalisations.[11]

CotB2 F149L

For the characterization of the CotB2 F149L product, a recently reported preliminary terpene production strain was used.[196] In contrast to the aforementioned *E.coli* strain, the engineered strain was simplified, missing the overexpressed genes IspF and IspD. Instead, the production system consisted of the *E.coli* Bl21(DE3) strain harboring the plasmids pCDFDuet-1 (*idi*, *dxs*), pACYCDuet-1 (*crTE*) and pet24a (*cotb2 F149L*). A further change was the cultivation in LB medium at 16°C for 60 hours in absence of C18 reversed phase beads. Consequently, the production yields were significantly reduced yielding 2.5 mg of the purified CotB2 F149L cyclisation product from a 6 liter shake flask cultivation.

Table 4. NMR data of cyclooctat-3-en-7-ol.

#	δ_H (ppm), *J* (Hz)	δ_C (ppm)		#	δ_H (ppm), *J* (Hz)	δ_C (ppm)	
1	a 1.49 (m, 1H) b 1.22 (m, 1H)	36.34	CH_2	**11**	-	43.60	C
2	1.78 (m, 2H)	53.86	CH_2	**12**	1.34 (m, 2H)	43.37	CH_2
3	-	80.96	C	**13**	a 1.57 (m, 1H) b 1.44 (m, 1H)	24.02	CH_2
4	a 1.94 (m, 1H) b 1.84 (m, 1H)	41.71	CH_2	**14**	2.11 (m, 1H)	47.74	CH
5	a 1.94 (m, 1H) b 1.70 (m, 1H)	25.92	CH_2	**15**	1.82 (m, 1H)	29.29	CH
6	3.18 (t, *J* = 8.3, 1H)	42.12	CH	**16**	0.88 (d, *J* = 6.7, 3H)	23.71	CH_3
7	-	136.28	C	**17**	0.78 (d, *J* = 6.8, 3H)	18.98	CH_3
8	5.53 (t, *J* = 8.2, 1H)	129.20	CH	**18**	1.27 (s, 3H)	29.14	CH_3
9	2.02 (m, 2H)	23.75	CH_2	**19**	1.80 (s, 3H)	21.62	CH_3
10	1.58 (m, 1H)	55.87	CH	**20**	0.87 (s, 3H)	18.88	CH_3

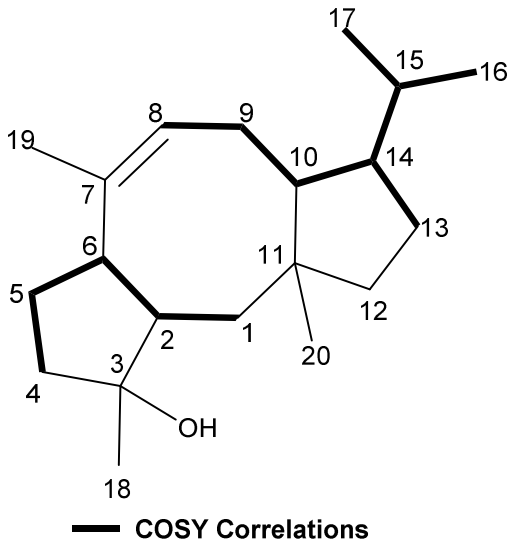


Figure 31. Structure of cyclooctat-3-en-7-ol. Key correlations in the COSY NMR spectrum are indicated.

^{13}C, DEPT 90/135 NMR spectroscopy revealed the presence of a single alkene bond, consisting of one tertiary and one quaternary olefinic carbon signal. A quaternary carbon signal at δ=80.96 (Table 4) displayed the characteristic chemical shift of a ternary alcohol function, which was in accordance with the observed molecular mass of 290 Da corresponding to $C_{20}H_{34}O$. The presence of four quaternary carbon signals were particularly noticeable (Table 4). This number

was equivalent to the wild type structure cyclooctat-9-en-7-ol and indicated the presence of a three ring structure. Analysis of the ^{1}H NMR spectrum showed an unusual signal at δ=3.18 ppm (t, ^{3}J(H,H)=8.3 Hz, 1H, CH) with triplet multiplicity (Table 4). Analysis of this signal by COSY NMR spectroscopy revealed correlations between the proton signals at δ=1.94 (Ha-5), 1.78 (H-2) and 1.70 (Hb-5) ppm and indicated the presence of a pseudo triplet. The connectivity of the carbon skeleton was deduced by COSY, HSQC and HMBC NMR spectroscopy (Appendix Figure 31-36) and is displayed in figure 31. Two spin coupling systems between H_2-1/H-2, H-2/H-6, H-6/H_2-5, H_2-5/ H_2-4 and H-8/H_2-9, H_2-9/H-10, H-10/H-14, H-14/H_2-13, H-14/H-15, H-15/H3-16, H-15/H_3-17 were detected in the COSY NMR spectrum (Figure 31). Using the HMBC key correlations H-2/C-3, H_3-18/C-2, H_3-18/C-3, H_3-18/C-4, H-6/C-3, H-2/C-3, H-5/C-3, H-6/C-3 the first spin system was assigned to a 5 membered ring containing a ternary alcohol. Bringing the second spin system into structural context was challenging, in particular since for H_2-12 no correlation was found in the COSY spectrum. The latter was assigned by the HMBC correlations H_2-12 /C-10, H_2-12/C-14 and H_2-12/C-20 in proximity to the second spin system. Spin coupling between H_3-17/H-15 and H_3-16/H-15 showed the presence of an isopropyl group analog to the cyclooctat-9-en-7-ol. The isopropyl group was connected to C-14 using the HMBC correlations H-10/C-11, H_2-13/C-11, H_2-12/C-11, H_2-13/C-15, H-15/C-10 and H-15/C-13. Both 5 ring structures were found to be connected to a 8 membered ring structure according to the HMBC key correlations H_2-1/C-11, H_2-1/C-10, H_2-1/C-12, H_2-C9/C-10, H_2-C97C-11, H-8/C-10 and H_2-1/C-3, H-6/C-7, H-6/C-19, H-6/C-8, H-6/C-9. The methyl group H_3-18 was connected to C-3 by the HMBC correlation with C-3, C-2 and C-4. H_3-19 was connected to C-7 by the correlations with C-6, C-7 and C-8. H_3-20 was connected to C-11 after displaying correlations with C-1, C-10 and C-11. Interestingly, structure elucidation revealed the production of a fusicoccane like diterpene. While the core 5-8-5 ring motif of the parental compound is maintained, the double bond (from C-9/10 to C-7/8) and the hydroxyl group (from C-7 to C-3) have migrated. No structure database (Reaxys)[193] entry match was found and the novel non-natural fusicoccane diterpene was named cyclooctat-3-en-7-ol.

CotB2 F107Y-B

GC-MS based screening of the CotB2 F107Y mutant displayed two products referred to as F107Y-A and F107Y-B. Unfortunately, purification of the F107Y_A product failed and was not analyzed by NMR spectroscopy. By constrast, a 6 L shake flask production (60 hours at 16°C)

using the aforementioned simplified *E.coli* system yielded about 1 mg purified F107Y-B.[196] MS spectrometry data of the F107Y_A/B products showed a molecular weight of 272 Da pointing out the absence of a hydroxyl group**.**

Table 5. NMR data of cyclooctat-1,7-diene.

#	δ_H (ppm), *J* (Hz)	δ_C (ppm)		#	δ_H (ppm), *J* (Hz)	δ_C (ppm)	
1	5.46 (s, 1H)	132.20	CH	**11**	-	47.05	C
2	-	153.99	C	**12**	1.57 (m, 2H)	43.20	CH_2
3	2.42 (m, 1H)	39.23	CH	**13**	a 1.48 (m, 1H) b 1.31 (m, 1H)	23.15	CH_2
4	a 1.88 (m, 1H) b 1.38 (m, 1H)	34.08	CH_2	**14**	2.16 (m, 1H)	47.45	CH
5	1.63 (m, 2H)	29.16	CH_2	**15**	1.86 (m, 1H)	28.37	CH
6	3.74 (d, *J* = 7.9, 1H)	42.33	CH	**16**	0.88 (d, *J* = 6.8, 3H)	23.47	CH_3
7	-	141.85	C	**17**	0.78 (d, *J* = 6.8, 3H)	18.98	CH_3
8	5.43 (t, *J* = 8.1, 1H)	128.04	CH	**18**	1.08 (d, *J* = 7.1, 3H)	22.10	CH_3
9	α 2.38 (m, 1H) β 2.03 (m, 1H)	23.63	CH_2	**19**	1.72 (s, 3H)	20.63	CH_3
10	1.71 (m, 1H)	50.21	CH	**20**	1.06 (s, 3H)	18.98	CH_3

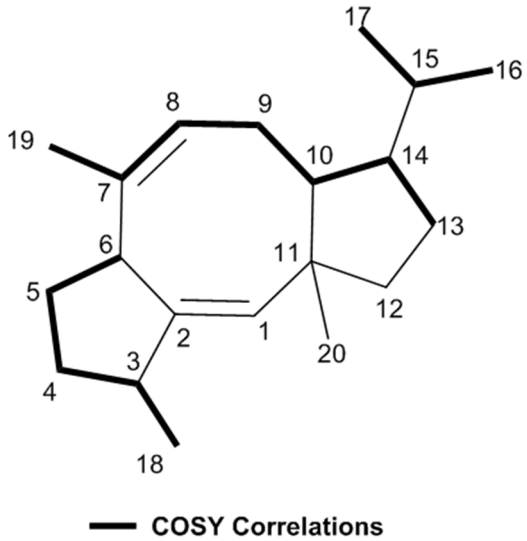


Figure 32. Structure of cyclooctat-1,7-diene. Key correlations in the COSY NMR spectrum are indicated.

^{13}C, DEPT 90/135 NMR spectroscopy (Appendix Figure 38) of the F107Y_B product revealed the existence of two olefinic bonds, one tertiary and one quaternary olefinic carbon signal (Table

5). The presence of four tertiary carbon atoms strongly suggested the presence of a three membered ring structure. ^{1}H NMR spectroscopy (Appendix Figure 37) showed signals for two olefinic protons at δ=5.46 (s, 1H, CH) and 5.43 ppm (t, ^{3}J(H,H)=8.1 Hz, 1H, CH) (Table 5). Interestingly, a signal at δ=3.74 ppm displayed pseudo-duplet multiplicity, which was elaborated by COSY NMR (Appendix Figure 39) and assignment to H-6 by the spin coupling with the two diastereotopic protons at δ=1.63 ppm (m, 2H, CH_2). Three spin systems were identified by analysis of the COSY spectrum. The first spin system comprised spin coupling between H_3-C-18/H-3, H-3/H_2-4, H_2-4/H_2-5, H_2-5/H-6, the second between H_3-19/H-8, H-8/H_2-9, H-9/H-10, H-10/H-14, H-13/H-14 and the third between H_3-17/H-15 and H_3-16/H-15 (Figure 32). Interestingly, proton coupling between H_3-19/H-8 in the second spin system represented a rare ^{4}J coupling. The last spin system was readily assigned to the characteristic terminal isopropyl group, which was in agrement with the observed doublet multiplicity of H_3-16 and H_3-17 observed in the ^{1}H NMR spectrum. The spin system created by the terminal isopropyl group was isolated. It was connected to C-14 by the HMBC correlations H-15/C-13 and H-15/C-10. The fist spin system was found to constitute a 5 membered ring by the HMBC correlations (Appendix Figure 41) H-3/C-2, H-3/C-4, H-3/C-5, H-3/C-6, H-6/C-2, H-6/C-3, H-6/C-4, H-6/C-5. This 5 membered ring could be linked to the second spin system by the HMBC correlations H-6/C-19, H-6/C-7 and H-6/C-8 indicating a connection between C-6 and C-7. Complementation of this carbon structure with the remaining carbon atoms is in accordance with the DEPT 90/135 assigned connectivity and quickly led to the assembly of a fusicoccane like macrocycle (Figure 32). The second five membered ring was assigned by the HMBC correlations between H_2-12/C-11, H_2-13/C-11, H-10/C-11. The assignment of the 8 membered ring was completed by the HMBC correlations H-1/C-11, H-1/C-2, H-1/C-12. The methyl group H_3-20, which showed no proton coupling was connected to C-11 by the HMBC correlations H_3-20/C-1, H_3-20/C-10 and H_3-20/C-11.

The new structure produced by CotB2 F107Y was searched for in the structure database (Reaxys)[193], but was undiscoverable. As a result, the new non-natural fusicoccane was named cyclooctat-1,7-diene. Compared to the wild type structure, cyclooctat-1,7-diene is characterized by the absence of a hydroxyl group and the presence of two alkene bonds between C-1/C-2 and C-7/C-8. Interestingly, the alkene bond between C-7/C-8 is also observed in cyclooctat-3-en-7-ol produced by CotB2 F149L.

To determine the relative stereochemistry of the compounds cyclooctat-7-en-3-ol and cyclooctat-1,7-diene, NOESY NMR spectrum were recorded. NOE enhanced correlations between protons were verified by molecular modelling using molecular mechanics (MM2) calculations. In order to match the observed ^{1}H NMR multiplicity for the proton H-6 in both compounds, structural solutions were evaluated which had dihedral angles (H-6 to H-5a) close to 90°C and matched the observed NOESY signals. Based on these data the stereochemistry for cyclooctat-7-en-3-ol was suggested and published as 2R*, 3S*, 6S*, 10S*, 11R*, 14S* and 3R*, 6S*, 10R* 11R*, 14S* for cyclooctat-1,7-diene (Figure 33).[197]

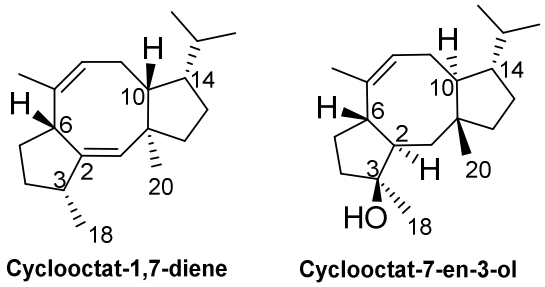


Figure 33. Proposed relative conformations of cyclooctat-1,7-diene and cyclooctat-7-en-3-ol.

Recently, a stereoselective cyclisation mechanism of cyclooctat-9-en-7-ol was suggested based on deuterium labeled GGDP precursors.[198] The cyclisation mechanism is in perfect agreement with the absolute configuration of R-cembrene A, (1R,3E,7E,11S,12S)-3,7,18-dolabellatriene and the carbon skeleton of cyclooctat-7-en-3-ol and cyclooctat-1,7-diene.[198] However, the previously determined relative configuration of cyclooctat-7-en-3-ol, cyclooctat-1,7-diene (Figure 33) was in conflict with the proposed mechanism. As a consequence, the NOESY NMR data (Appendix Figure 36, 42) from these substances was re-evaluated for this thesis.

3.3.5 Re-Determination of the Relative Configuration of Cyclooctat-7-en-3-ol and Cyclooctat-1,7-Diene

Molecular modelling of cyclooctat-7-en-3-ol and cyclooctat-1,7-diene was done using the most likely mechanism based configuration (see below Figure 34, Figure 35 and Figure 36) and validated based on the observed NOE correlation.

Cyclooctat-7-en-3-ol

The most notable structural element of cyclooctat-7-en-3-ol is the *cis*-configuration between the protons H-6 and H-2. The proton signal of H-6 (δ=3.18 (t, ^{3}J(H,H)=8.3Hz, 1H, CH)) displayed pseudo triplet multiplicity. However, spin coupling analysis revealed coupling with H-5a, H-5b and H-2, which usually results in a multiplet multiplicity. An explanation for this different multiplicity can found in the karplus relationship. According to the $^{3}J_{H,H}$ karplus relationships, the $^{3}J_{H,H}$ coupling constant is reduced to zero once the dihedral angle between two protons reaches 90°C.[199] Consequently, a pseudo triplet by three spin coupling partners requires one dihedral angle to be 90°C and the remaining dihedral angle to account for equivalent $^{3}J_{H,H}$ coupling constants. The dihedral angles in the mechanism based modelled cyclooctat-7-en-3-ol structure were H-6/H-5a = -39°, H-6/H-5b = -79° and H-6/H-2 = 35. While the corresponding angles of -39° and 35° account for comparable $^{3}J_{H,H}$ coupling constants, the angle of -79° only approximately fulfils the requirements for pseudo triplet multiplicity. NOESY NMR spectroscopy revealed NOE correlations between H-6/H-2, which were in agreement with an anticipated *cis*-configuration. Correlations between H-6/H_3-20, H-2/H_3-20 suggested that the methyl group H_3-20, H-6 and H-2 were *cis*-oriented to each other, which was also in accordance with the model. Furthermore, the observed correlations between H_3-18 and H_3-19 indicated a *cis*-configuration between both methyl groups and a *trans*-configuration between both methyl groups to H-6 and H-2, respectively. The configuration of the isopropyl group was assigned as β, according to the NOE correlation between H_2-9 and H-15. The observed NOE correlations as well as the reported dihedral angles did not contradict the mechanism derived absolute configuration of cyclooctat-7-en-3-ol as 2S, 3S, 6S, 11R, 14R.

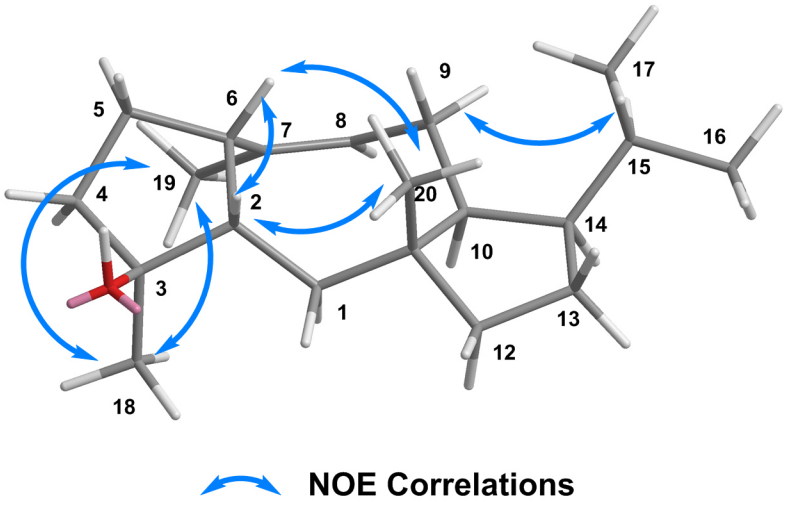


Figure 34. Molecular model (MM2) of (2S, 3S, 6S, 11R,14R) cyclooctat-7-en-3-ol. Key NOESY correlations are illustrated.

Cyclooctat-1,7-diene

The mechanism derived configuration of cyclooctat-1,7-diene is very related to the previous suggested structure. The stereocenters 10R*, 11R* and 14S* are in agreement with the suggested absolute conformation. However, the stereocenters 3R and 6R in the structural model were previously assigned as S stereocenters.

The ^{1}H NMR spectrum of cyclooctat-1,7-diene showed the presence of a pseudo duplet at δ=3.32 ppm (d, ^{3}J(H,H)=7.9Hz, 1H, CH) for H-6. According to the $^{3}J_{H,H}$ karplus relationships[199] a pseudo duplet requires the dihedral angles between H-6/H-5a or H-6/H-5b to be 90°. Measuring of the dihedral angles in model structure of cyclooctat-1,7-diene resulted in H-6/H-5a = -36° and H-6/H-5a = -83°. The latter angle is in good proximity to 90° and would result in a $^{3}J_{H,H}$ coupling constant close to zero. NOESY NMR spectroscopy revealed NOE correlations between H-6/H_3-20 and H-6/H-9α indicated a close proximity between H-6, H_3-20 and H-9α, respectively. Accordingly, the methyl group H_3-20 and H-6 are in β-configuration. Furthermore, correlation between H-9β and H-15 suggested a β-configuration for the isopropyl group. A correlation between H-8 and H-10 proposed a α-configuration of H-10. Correlation between H-6 and H-3 indicated a *trans*-configuration between H-6 and the methyl group H_3-18, which was further confirmed by correlation between the methyl groups H_3-18/ H_3-19. These observed NOE correlations are consistent with the mechanism based absolute configuration 3R, 6R, 10S, 11R, 14S.

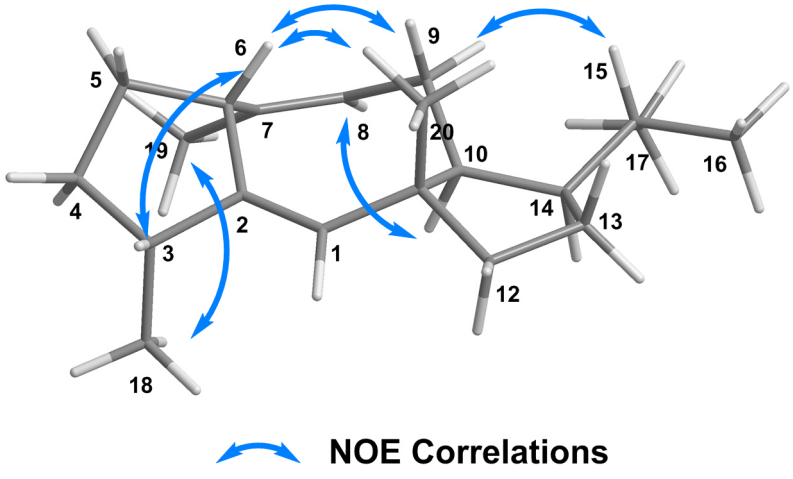


Figure 35. Molecular model (MM2) of (3R, 6R, 10S, 11R, 14S) cyclooctat-1,7-diene calculated by MM2. Key NOESY correlations are illustrated.

3.3.6 Mechanistic Consideration for the Production of CotB2 Mutants

The recently proposed stereoselective cyclisation mechanism of cyclooctat-9-en-7 is illustrated on figure 36.[198]

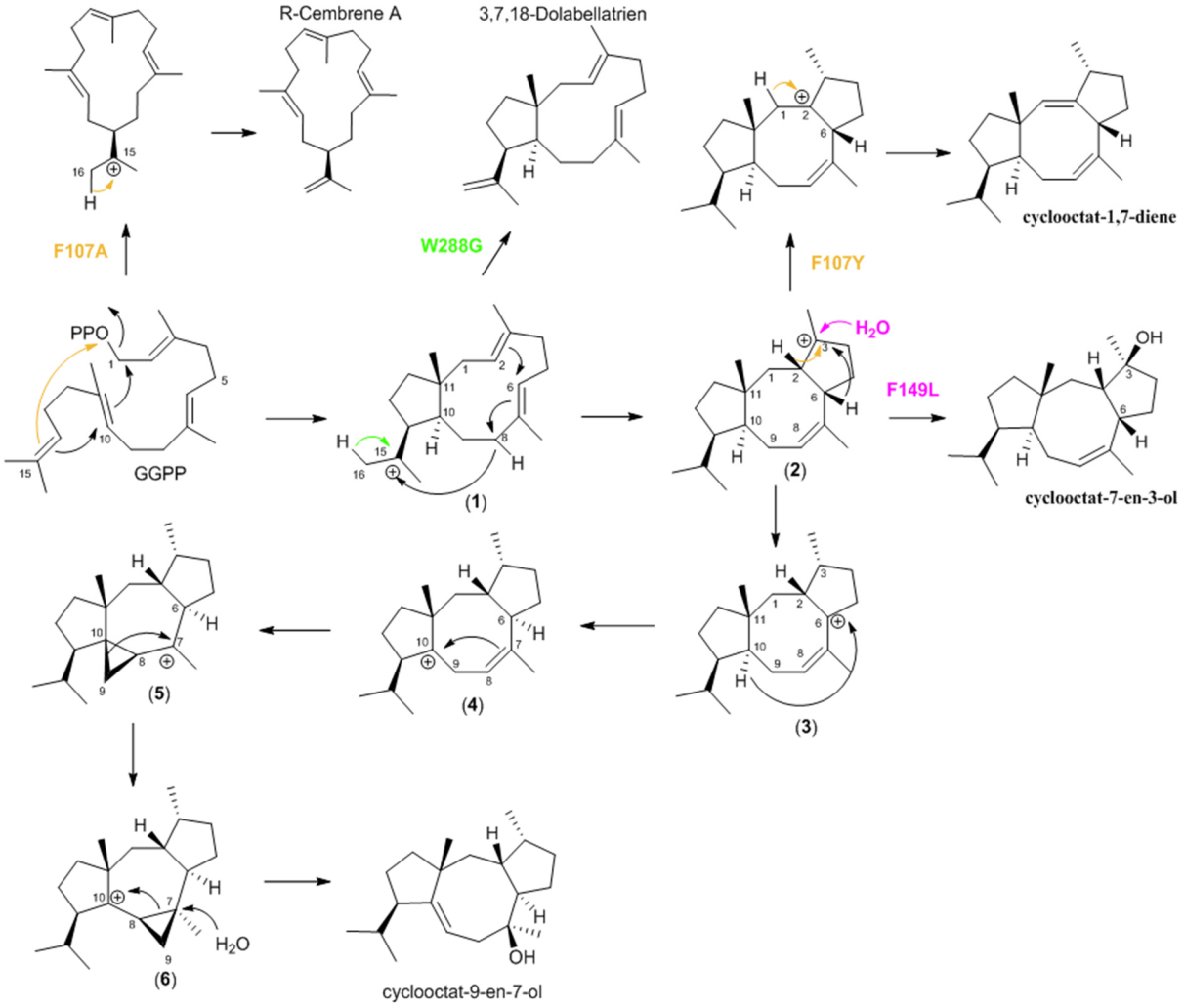


Figure 36. Proposed cyclisation of the wild type cyclooctat-9-en-7-ol[198] and the CotB2 mutant products. CotB2 Phe149 in pink, CotB2 Phe107 in orange and CotB2Trp288 in green.

Cyclisation of GGDP to cyclooctat-9-en-7-ol is more complicated than anticipated and includes two 1,5 hydride shifts and the formation of a cyclopropyl intermediate ring structure (see Figure 36).[198] After cleavage of the pyrophosphate group the bicyclic intermediate (**1**) is formed, harboring the carbocation at C-15.[198] In the next step a 1,5 hydride shift transfers the proton from C-8 to C-15 generating the allylic cation at C-8.[198] The cation is shifted to C-6, where a nucleophilic attack by the double bond between C-2/C-3 forms the tricyclic 5-8-5 ring structure (**2**).[198] Subsequently, the newly generated carbocation at C-3 is transferred to C-6 by a 1,3 hydride shift creating the intermediate (**3**).[198] The second 1,5 hydride shift from C-10 to C-6

transfers the carbocation to C-10 yielding intermediate (**4**).[198] The latter undergoes structural rearrangement forming a cyclopropyl ring comprising C-8/C-9/C-10 in the intermediate (**5**), which is created by a nucleophilic attack of the double bond between C-7/C-8 on the carbocation at C-10.[198] The cyclopropyl ring undergoes further rearrangement resulting in intermediate (**6**).[198] A nucleophilic attack of a water molecule on the cyclopropyl ring at C-7 eventually converts intermediate (**6**) to the final product cyclooctat-9-en-7-ol.[198]

The mutant CotB2 products 3,7,18 dolabellatriene, cyclooctat-7-en-3-ol and cyclooctat-1,7-diene are consistent with the proposed mechanism. In contrast, the diterpene R-cembrene A created by CotB2 F107A does not reflect the proposed cyclisation mechanism for cyclooctat-9-en-7-ol. Although the cyclisation mechanism for R-cembrene A is currently unknown, it can be assumed that the cyclisation is initiated by the electrophilic attack of the carbocation at C-1 on the double bond between C-14/C-15. The resulting carbocation at C-15 is then quenched by proton elimination from C-16. Interestingly, the configuration of the isopropenyl group in R-cembrene A is in the β form, which is similar to the wild type structure cyclooctat-9-en-7-ol and all other generated mutant structures. This finding suggests the GGDP substrate in the enzyme pocket only partially adopts a new orientation. This orientation enables a sterically preferred electrophilic attack of the carbocation on the *si face* of C-14 in the the double bond between C-14/C-15 compared to a *si face* attack at C-10 in the double bond between C-10/C-11 that is part of the natural cyclisation mechanism.

An explanation for the production of 3,7,18 dolabellatriene by CotB2 W288G is the early quenching of the intermediate (**1**). The stereochemistry of the isopropyl group, H-10 and H_3-20 complies with the cyclisation mechanism and stays fixed for the remaining cyclisation intermediates. This stereochemistry is conserved in the wild type product and the non-native structures cyclooctat-7-en-3-ol and cyclooctat-1,7-diene. Hence, this finding indicates that the orientation of GGDP is left unaltered by the W288G mutation. The amino acid W288 may be essential to stabilize the carbocation at C-15 in the intermediate (**1**). Consequently, the mutation W228G suppresses the 1,5 hydride shift that follows the formation of intermediate (**2**). Instead, the carbocation is quenched by proton abstraction at C-15 yielding the dolabellatriene macrocycle.

The production of cyclooctat-7-en-9-ol by CotB2 F149L is best explained by quenching of intermediate (**2**) with water. The protons H-2 and H-6 in cyclooctat-7-en-9-ol are in β-configuration. The proposed mechanism suggests the abstraction of H-6 by a suprafacial 1,3-

hydride shift to yield intermediate (**3**). However, this hydride shift is suppressed by the F149L mutant. As a result, the carbocation at C-3 is attacked by a water molecule. Stereochemical considerations deduce the nucleophilic attack to occur on the identical side of the molecule, where the suprafacial 1,3-hydride shift of the wild type mechanism is happening.

Cyclooctat-1,7-diene derived from the CotB2 F107L mutant is also well explained by using intermediate (**2**) as a starting point. In contrast to the CotB2 F149L mutant, the carbocation at C-3 is not quenched by water addition. Instead, it is likely that a new 1,2 suprafacial hydrid shift from C-2 to C-3 takes place and translocates the positive charge to C-2. The carbocation is eventually neutralized by proton elimination at C-1, which led to the formation of a double bond between C-1 and C-2. Interestingly, instead of directly facilitating proton elimination at C-4 to yield the formation of a new double bond within the 5 membered ring, the formation of a double bond within the 8 membered ring is preferred. This finding may be explained by avoidance of unfavorable ring strains.

3.3.7 Structural Consideration for the Production of CotB2 Mutants

The data from the *in silico* generated closed complex of CotB2 and molecular dynamic simulation performed with the GGDP substrate was obtained from the publication: "Identification of amino acid networks governing catalysis in the closed complex of class I terpene synthases." Schrepfer, P.; Buettner, A.; Goerner, C.; Hertel, M.; van Rijn, J.; Wallrapp, F.; Eisenreich, W.; Sieber, V.; Kourist, R.; Brück, T. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, E958–E967.

Comparison of the Closed (*in silico*) and Open Complex of CotB2

The crystal structure of CotB2 presents the open and therefore catalytically inactive enzyme conformation. Additionally, the structure is missing Mg^{2+} ions and a substrate or analogue in the active site. Although this data provides some valuable insights into the structural organizing of the active site, it can only provide limited information on the amino acid residues involved in the dynamic carbocation migration processes of the cyclisation reaction. To put the cyclisation mechanism of the wild type enzyme and the mutants into the context of the active site, a suitable model was created. To date a catalytically relevant terpene synthase in the closed complex containing the native GGPP substrate remains elusive. Consequently, the recently published closed complex of the sesquiterpene selinadiene synthase in complex with the substrate analog dihydrofarnesyl pyrophosphate and three Mg^{2+} ions was used as structural

template.[180] Interestingly, this enzyme is structurally closely related to the open CotB2 complex displaying a r.m.s.d of 2.171 Å over 144 residues with 17.36% primary sequence identity. To create the productive closed complex model of CotB2, the amino acids from the selinadiene synthase were mutated to the corresponding CotB2 counterpart as shown by Schrepfer et al.[181] To determine the orientation of the GGDP substrate in the closed complex model of CotB2, the substrate analog was replaced *in silico* with GGDP followed by energy minimization and MD simulation.[181] Furthermore, molecular dynamic simulations were carried out to determine the orientation of the reaction intermediates 1 -6 in the closed complex model.[181]

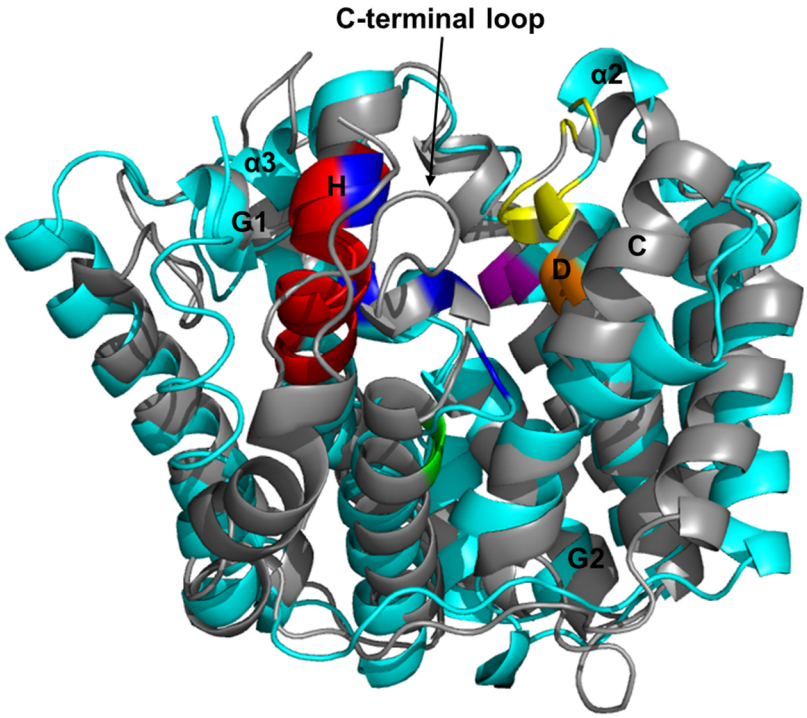


Figure 37. Superposition of the open complex crystal structure of CotB2 (PDB: 4OMG) and the *in silico* generated closed complex of CotB2. The open complex of CotB2 is shown in cyan and the closed complex is shown in grey. Important helices, which display structural differences are entitled. The aspartate amino acids of the DXDD motif are depicted in yellow. The NSE/DTE motif is depicted in red. Arginine amino acids presumably involved in the Mg^{2+} and substrate binding are depicted in blue. Amino acids involved in production of altered diterpenes are indicated: Phe149 in purple, Phe107 in orange and Trp288 in green.

Figure 37 illustrates an alignment of the *in silico* calculated closed complex (grey) and the open CotB2 (cyan). The alignment showed a r.m.s.d of 2.130 Å over 190 residues with 100%. The center of the active site harbors the catalytically important amino acids F107 (orange), F149 (purple) and W288 (green), which are structurally almost unaltered. Interestingly, the helices G1 and G2 as well as the in-between loop structure harboring the helix break motif (^{177}RVTDI^{181}) show an almost identical fold in the closed and open conformation. The break motif was reported to be essential for the cleavage of the pyrophosphate group by undergoing a structural rearrangement upon substrate binding (induced fit) to form the closed complex.[180] However, since an almost identical folding was observed in the open complex crystal structure

of CotB2, the loop structure might undergo further relocation within the real but experimentally inaccessible closed complex of CotB2.

Structural deviants in the alignment are observed at the entry of the active site, which harbors the essential amino acids for the Mg^{2+} complexation and binding of the substrate's pyrophosphate group. A minor variance is illustrated in helix D, helix α2 harboring the DDXD motif (yellow) and the connecting loop structure between both helices. All structural elements are shifted towards the active side of the protein causing Asp 110 and Asp 111 to protrude deeper into the active site. Additionally, the upper part of Helix H carrying the 'NSE/DTE' motif (red), which is involved in the Mg^{2+} binding is leaning towards the active center causing the side chain of Asn 220 to point towards the DDXD motif. Since the open complex of CotB2 is missing essential parts of the C-terminal loop, structural amendment of this loop from selinadiene synthase caused Arg 294 to relocate and thereby pointing to Asp 110. Furthermore, the entry of the active site is partially covered by the translocated helix C and a loop structure, which was formally assigned to helix α3 in the open complex. Closure of the active side channel is in agreement with the proposed formation of a Michaelis complex, which was previously observed in the selinadiene synthase.[180]

Active Site of the *in silico* Modelled Closed Complex of CotB2

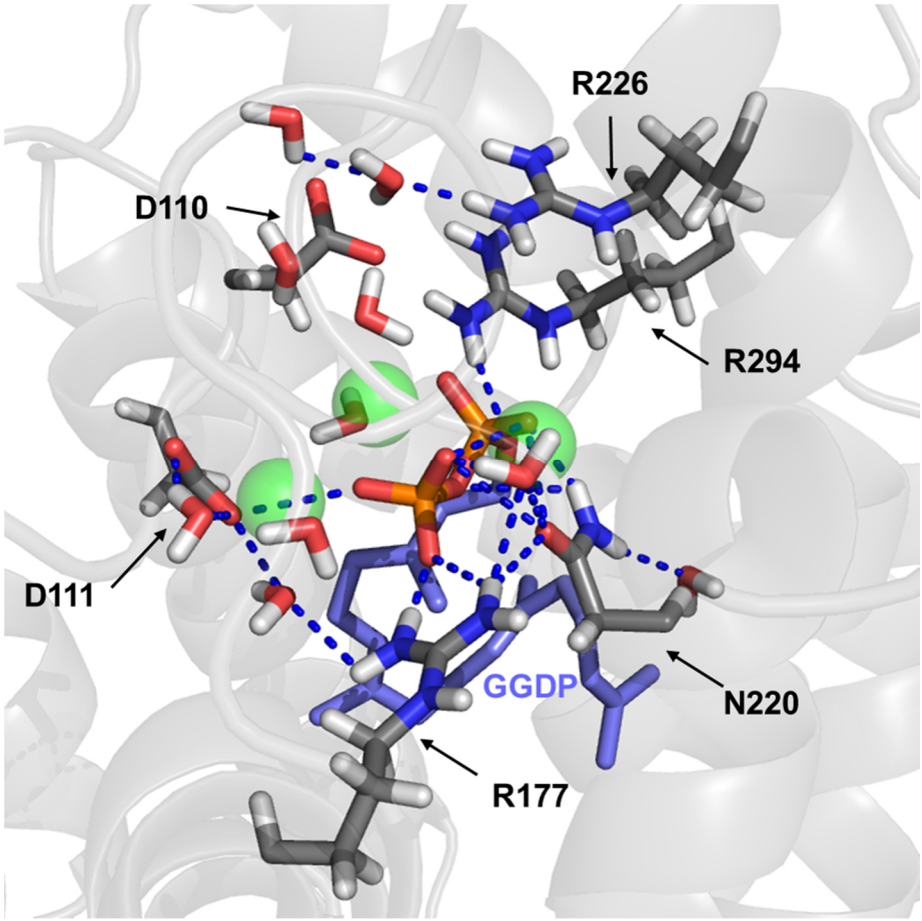


Figure 38. Catalytic active site of the *in silico* generated closed complex of CotB2 harboring the GGDP substrate, $(Mg^{2+})_3$ cluster and several water atoms. The carbon chain of the GGDP substrate in the "S" shape configuration is depicted in purple. The pyrophosphate is depicted in orange (phosphor) and red (oxygen). The sphere of the three Mg^{2+} ions forming ion bonds with the pyrophosphate group are depicted in green. Highlighted amino acid side chains are depicted in grey, corresponding nitrogen atoms are coloured in blue, oxygen atoms in red and hydrogen atoms in white.

In the active site of the modelled closed complex, the $(Mg^{2+})_3$ cluster and the pyrophosphate group of GGDP are coordinated by several amino acids such as Arg 177, Asn 220 from the NSE/DTE motif, Asp 110, Asp 111 from the DDXD motif as well as Arg 226, Arg 294 and several water atoms (Figure 38). It clearly follows that Asp 113 from the DDXD motif is misplaced and not part of this coordinating network. Asn 220 and Arg 177 form hydrogen bonds by themselves, the pyrophosphate group and two water molecules. On the opposite site of the pyrophosphate group, a hydrogen bond network coordinating several water molecules is established between Asp 110 and Asp 111. These water molecules are part of the coordination network for the Mg^{2+} ions. While Arg 294 is forming a hydrogen bond with the pyrophosphate group, Arg 226 is positioning water molecules by means of Asp 111. The pyrophosphate group of the GGPP substrate is fixed by hydrogen bonds and electrostatic interactions with the $(Mg^{2+})_3$ cluster.

Orientation of the GGPP Substrate in the *In Silico* Modelled Closed Complex of CotB2

In the closed complex the GGPP substrate is oriented in an "S-shaped" conformation, which is in very good agreement with the proposed cyclisation mechanism (Figure 39). In this conformation C-1 is in close proximity to the *si face* of the double bond C-10/C-11. Furthermore, the double bond C-14/C-15 is in close proximity to the double bond (C-10/C-11). Consequently, GGDP is oriented in the catalytically active conformation. This substrate orientation already defines the beta conformation of the methyl group at C-11 and the isopropyl group at C-15 as well as the α-conformation of the H-10 proton, which is captured in the dolabellatriene intermediate.

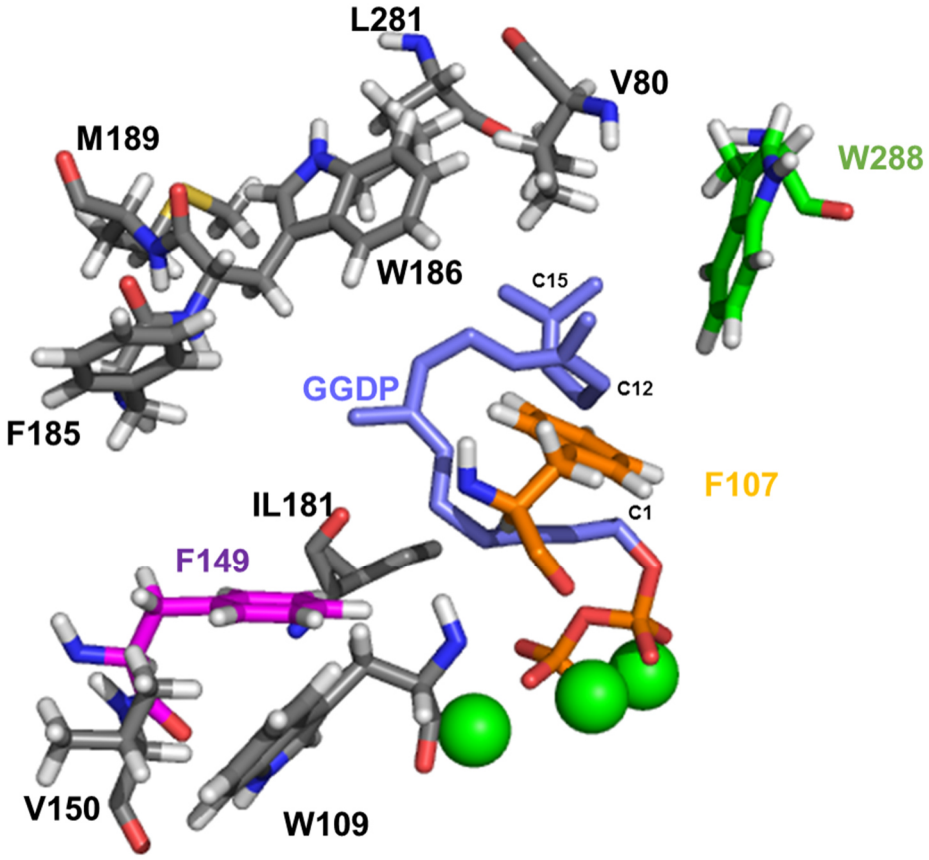


Figure 39. Selected residues in close proximity of the GGDP substrate from the *in silico* generated closed complex of CotB2. The carbon chain of the GGDP substrate in the "S" shape configuration is depicted in purple. The pyrophosphate is depicted in orange (phosphor) and red (oxygen). The spheres of the three Mg^{2+} ions forming ion bonds with the pyrophosphate group are depicted in green. Amino acid side chains are depicted in grey, corresponding nitrogen atoms are coloured in blue, oxygen atoms in red, hydrogen atoms in white and sulfur atoms in yellow. Trp288 is depicted in green and Phe149 in magenta.

The model reveals that the "S-shape" orientation of the substrate is surrounded by the hydrophobic amino acids Val 80, Phe107, Trp109, Phe149, Val150, Ile181, Phe185, Met189, Trp186, Leu281 and Trp288 (Figure 39). Interestingly, F107 is directly positioned within the 11 membered ring, which is formed in intermediate (**1**). In this study it was demonstrated that by

removing the phenylalanine side chain in the F107G mutant, the "S-shape" orientation is disintegrated to yield R-cembrene A. Hence, it is highly likely that the side chain of F107 is essential for the correct orientation of the GGDP substrate. Mechanistically, the formation of R-cembrene A is initiated by a *si face* attack of the carbocation (C-1) on the double bond between C-14/C-15. This reaction requires the dislocation of the double bond between C-11/C-12 in favour of the double bond between C-14/C-15 to facilitate close proximity to C-1.

Induced-Fit Mechanism in CotB2

The induced-fit model upon substrate binding was previously postulated for class I terpene synthases based on the closed complex of the selinadiene synthase.[180] Since the latter is structurally related to CotB2 and was used as template for modelling the closed complex, a comparison of the catalytic properties is intriguing. The selinadiene synthase catalyses the cyclisation of farnesyldiphosphate (FPP) to the sesquiterpene selina-4(15),7(11)-diene.[180] The crystal structure harboring the substrate analog has been reported in the open and closed conformation.[180] It was found that an effector triad consisting of Arg 178, Asp 181, and Gly 182, which is located in the loop structure between helix G1 and G2, is essential for the cleavage of the pyrophosphate group.[180] Upon binding of the FPP substrate, this loop is shifted towards the pyrophosphate group, which is caused by Arg 178 acting as a pyrophosphate sensor and by Asp 181 acting as a linker to mediate the rearrangement.[180] Due to this shift, the carbonyl oxygen atom of Gly182 comes in close proximity to the p* molecular orbital of the double bond between C-2/C-3 and thereby initiates the pyrophosphate cleavage.[180] An equivalent effector tirade is found in CotB2 comprising Arg 177, Asp 180 and Ile 181 (Figure 40). In analogy to the selinadiene synthase, Arg 177 in the closed complex of CotB2 is participating in the binding of the pyrophosphate group. Furthermore, Asp 180 is in close proximity to the pyrophosphate group and may indeed function as a linker. However, in contrast to the structure of the selinadiene synthase, neither the carbonyl group of Ile 181 nor the carbonyl group of any other amino acid from the loop structure is in sufficient proximity to the double bond between C-2/C-3 (see Figure 40). This finding demonstrates the limits of the *in silico* created closed complex, which is a good approximation but does not fully represent the real

complex. Nevertheless, the model allows the prediction of mechanistic details which are discussed below.

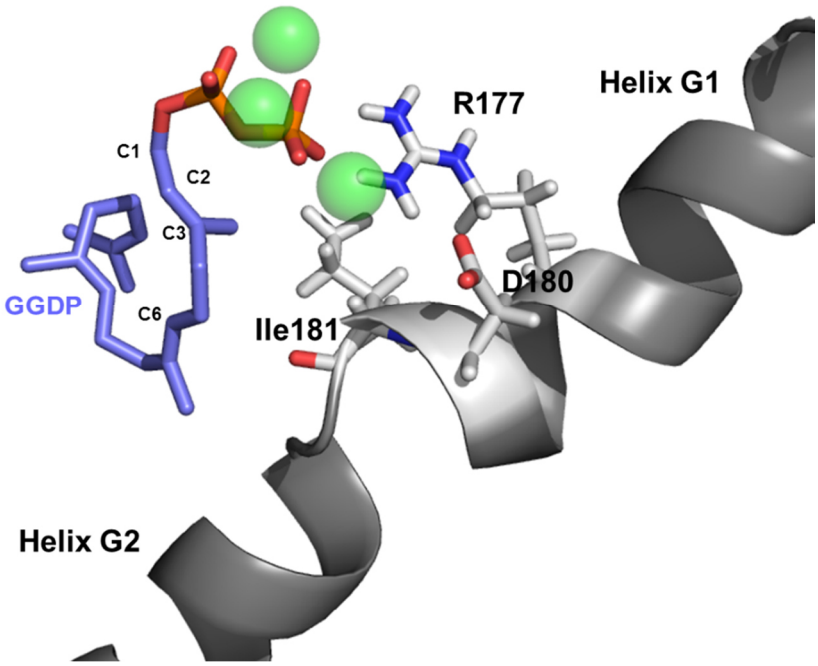


Figure 40. Cartoon representation of the G1 and G2 helices and the connecting loop structure derived from the *in silico* generated closed complex of CotB2 (grey) harboring the GGDP substrate (purple). The pyrophosphate group of GGDP substrate is depicted in orange (phosphor) and red (oxygen). The three Mg^{2+} ions forming ion bonds with the pyrophosphate group are depicted in green. Amino acid side chains forming the catalytic tirade are depicted in grey. Corresponding nitrogen atoms are coloured in blue, oxygen atoms in red and hydrogen atoms in white. Additionally, the carbonyl and amino group of the main chain from Ile 181 is illustrated.

Amino Acids Essential for the Formation of 3,7,18 Dolabellatriene

The amino acid W288, whose mutation caused the formation of the dolabellatriene diterpene, is located at the terminal isopropyl group of the GGPP substrate (Figure 41). Mechanistical considerations have allocated this amino acid to facilitate the 1,5 hydride shift from C-8 to C-15 after formation of intermediate (**1**). In the model harboring intermediate (**1**), is can be seen that C-8 and C-15 are in very close proximity making a 1,5 hydride shift feasible. Furthermore, the model suggests that W288 is not in the immediate vicinity to C-15 but rather to C-16. Consequently, its function may be to facilitate the correct orientation of the isopropyl group for the 1,5 hydride shift. Interestingly, Asn 284 is in close proximity to the isopropyl group and may catalyze the deprotonation reaction to yield dolabellatriene.

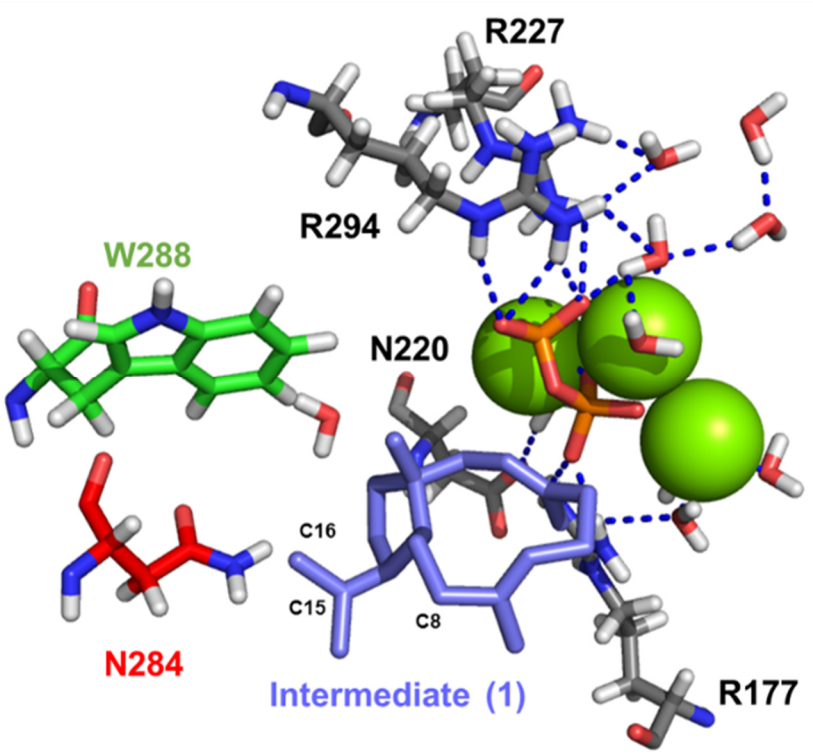


Figure 41. Selected residues in proximity of intermediate (**1**) in the active site of the *in silico* generated closed complex of CotB2. The carbon chain of intermediate (**1**) is depicted in purple. The pyrophosphate is depicted in orange (phosphor) and red (oxygen) The sphere of the three Mg^{2+} ions forming ion bonds with the cleaved pyrophosphate group are depicted in green. Amino acid side chains are depicted in grey. Corresponding nitrogen atoms are coloured in blue, oxygen atoms in red and hydrogen atoms in white. Trp288 is depicted in green and Asn284 in red.

Amino Acids Essential for the Formation of Cyclooctat-7-en-9-ol

After a 1,5 hydride shift intermediate (**2**) is formed, which is the structural basis for the production of cyclooctat-7-en-9-ol and cyclooctat-1,7-diene. For the production of cyclooctat-7-en-9-ol, this study showed that the mutation F149L plays an important role in quenching the carbocation at C-3 by water.

The model harboring intermediate (**2**) shows the amino acid F149 in close proximity to C-6 (Figure 42). Consequently, F149 in the wild type enzyme may be involved in stabilizing the carbocation at C-6 prior to the 1,5 hydride-shift form C-10 to C-6. Interestingly, a series of F149 mutants were shown to yield the same product including glycine, valine, leucine and histidine. These amino acids that display shorter hydrophobic side chains may be incapable of stabilizing the carbocation at C-6. Furthermore, these side chains may not be suitable to facilitate the correct orientation that is needed for the 1,5 hydride shift. As a result, the carbocation in the intermediate (**2**) remains at C-3 and is eventually quenched by water, rather than undergoing a 1,3 hydride shift to C-6. The model reveals the presence of several water molecules in close proximity to the *si face* of C-3, which are coordinated in the surroundings of the cleaved

pyrophosphate group. The pyrophosphate group may also provide the corresponding base to capture the proton. A *si face* nucleophilic attack is in excellent agreement with the stereochemistry of the tertiary alcohol function in cyclooctat-7-en-9-ol.

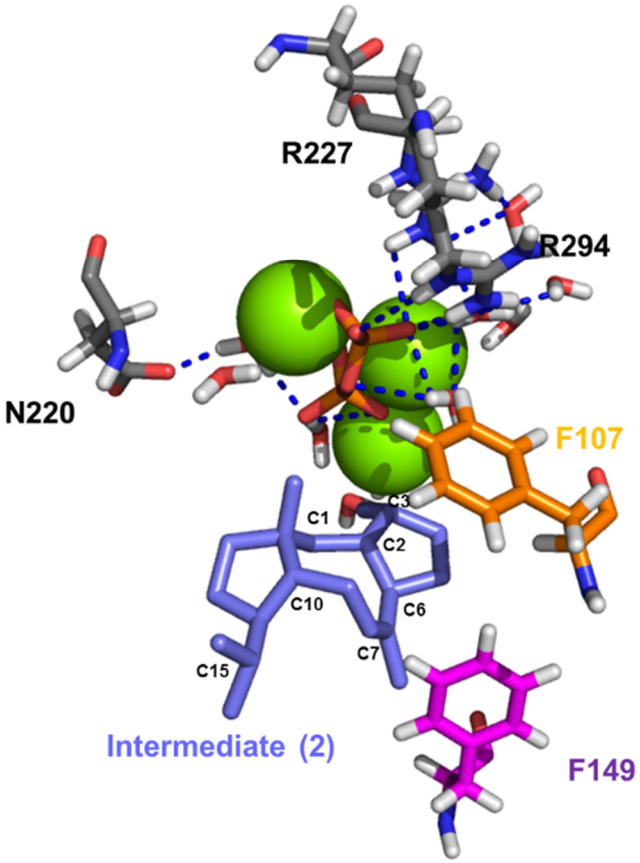


Figure 42. Selected residues in proximity of intermediate (**2**) in the active site of the *in silico* generated closed complex of CotB2. The carbon chain of intermediate (**2**) is depicted in purple. The pyrophosphate is depicted in orange (phosphor) and red (oxygen). The sphere of the three Mg^{2+} ions forming ion bonds with the cleaved pyrophosphate group are depicted in green. Amino acid side chains are depicted in grey. Corresponding nitrogen atoms are coloured in blue, oxygen atoms in red and hydrogen atoms in white. Phe107 is depicted in orange and Phe149 in magenta.

Amino Acids Essential for the Formation of Cyclooctat-1,7-diene

The formation of cyclooctat-1,7-diene and a second unidentified product is caused by the mutation F107Y. According to the model harboring intermediate (**2**), F107 is located above the five membered ring in close proximity to C-2 and C-6 (Figure 42). Mechanistically the production of cyclooctat-1,7-diene starts from intermediate (**2**), which undergoes two subsequent 1,2 hydride shifts from C-2/C-3 and C-1/C-2, followed by proton elimination at C-1. Since the side chains of phenylalanine and tyrosine are of comparable size, steric effects can be ignored. Instead, the reversed polarity caused by the introduction of a hydroxyl group needs to be considered. By positioning a negative charge above the five membered ring, the water

surrogate between C-3 and the pyrophosphate group is likely disturbed. This complicates a nucleophilic attack of water that would ultimately lead to the formation of cyclooctat-7-en-9-ol.

Furthermore, the hydride shift (C-3/C-6) in the wild type enzyme might be prevented by steric repulsion of the negative charges from the hydride anion and the lone pairs of the hydroxyl group. As a result the carbocation at C-3 undergoes a different reaction pathway. Interestingly, a series of F107 mutations such as leucine, isoleucine, serine and aspartic acid led to product mixtures, which could not be characterized. This might be an additional indication that the carbocation in intermediate (**2**) can be stabilized by proceeding to a series of different reaction pathways.

Origin of the Hydroxylgroup in Cyclooctat-9-en-7-ol

After the 1,3 hydride shift (C-3/C-6) intermediate (**3**) is formed in the wild type enzyme. The model confirms that intermediate (**3**) is oriented in a catalytic active conformation for a 1,5 hydride shift (C-6/C-10) (data not shown) following the formation of intermediate (**4**). Next, intermediate (**4**) undergoes structural rearrangement to intermediate (**5**) and subsequently to intermediate (**6**). In the last cyclisation step, the carbocation at C-7 located in the cyclopropane ring is quenched by water. The model shows a close proximity between Asn 103 and the carbocation at C-7. It is conceivable that Asn 103 plays a role in facilitating the nucleophilic attack of the water molecule on the carbocation at C-7 (Figure 43). Stereochemical consideration of the final product cycooctat-9-en-7-ol allocate the water molecule from the direction of the pyrophosphate group.

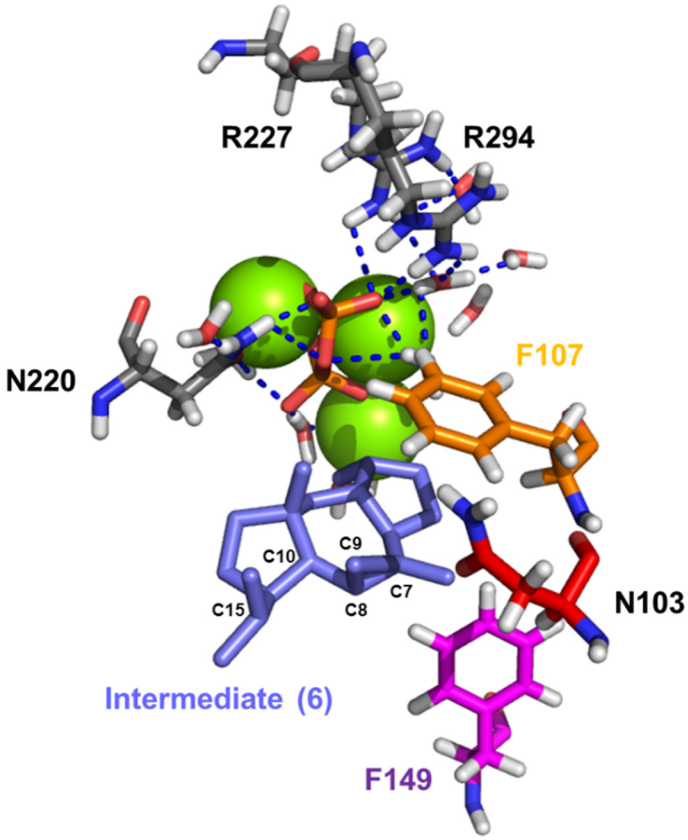


Figure 43. Selected residues in proximity of intermediate (**6**) in the *in silico* generated closed complex of CotB2. The carbon chain of intermediate (**6**) is depicted in purple. The pyrophosphate is depicted in orange (phosphor) and red (oxygen). The sphere of the three Mg^{2+} ions forming ion bonds with the cleaved pyrophosphate group are depicted in green. Amino acid side chains are depicted in grey corresponding nitrogen atoms are coloured in blue, oxygen atoms in red and hydrogen atoms in white. Phe107 is depicted in orange, Phe149 in magenta and Asn 103 in red.

3.4 Conclusion and Outlook for the CotB2 Mutagenesis Studies

In this study mutagenesis engineering was applied to the diterpene synthase Cotb2 to produce a series of diterpene macrocycles different to the wild type cyclooctat-9-en-7-ol. Therefore an *in silico* model in combination with crystallographic studies was used to identify several key catalytic active amino acids. Gene libraries for single amino acids were screened using an *in vitro* and *in vivo* screening approach. In the course of this study, it was possible to identify four enzyme mutants that were able to produce diterpene macrocycles different from the wild type.

The mutant CotB2 F107A resulted in the production of the single natural product R-cembrene A. This is the first report of the selective production of R-cembrene A from a diterpene synthase. To date only a bacterial synthase has been described that produces a mixture of cembrenes including R-cembrene A.[185] Cembrene terpenoids display a series of biological functions such as cytotoxic, pheromonal and antimicrobial activities.[200] Therefore, R-cembrene A is a good starting point for the development of novel bioactives. In this study the biotechnological production of R-cembrene A was affected by product degradation. This obstacle was effectually circumvented by applying hydrophobic beads to the cultivation medium. Using this strategy in combination with strain optimization techniques such as balanced promotors and genomic pathway integration will pave the way to increase the production titers.

The production of (1R,3E,7E,11S,12S)-3,7,18-dolabellatriene by CotB2 W288G resembles an important discovery. This structure displays antibiotic activity against multidrug-resistant *Staphylococcus aureus* without the need for further chemical modifications.[11] The dolabellane family is a rich source of biological active compounds encompassing antiviral, antibacterial, cytotoxic, molluscicidal and phytotoxic activity.[195] However, the lack of corresponding terpene synthases has limited the access to sufficient material for detailed structure function relationship studies. In this study several milligrams of (1R,3E,7E,11S,12S)-3,7,18-dolabellatriene were produced in *E.coli*. This is the first recombinant produced dolabellane terpene to date. Interestingly, no toxic effect was observed during the production of this antibiotic compound. As 1R,3E,7E,11S,12S)-3,7,18-dolabellatriene is presumably primarily effective against gram-positive bacteria such as *Staphylococcus aureus*, the gram-negative bacteria *E.coli* was insensitive to the amounts produced in this study. As a result, the production in *E.coli* resembles an efficient route to this potential pharmaceutically relevant compound.

Applying balanced promotors and genomic pathway integration will probably lead to a significantly improved production titer.

The production of the fusicoccin structures cyclooctat-7-en-3-ol and cyclooctat-1,7-diene was accomplished by the mutants CotB2 F149L and CotB2 F107Y. The mutant CotB2 F149L exclusively produced cyclooctat-7-en-3-ol, while the mutant CotB2 F107Y resulted in a mixture of two terpenes from which only one was characterized (F107Y-B). The recently reported cyclisation mechanism of the wild type CotB2 led to the re-evaluation and correction for the previously published relative stereochemistry of cyclooctat-7-en-3-ol and cyclooctat-1,7-diene.[198] Interestingly, the newly determined stereochemistry of (2S*, 3S*, 6S*, 11R*, 14R*) for cyclooctat-7-en-3-ol is identical to the sesterterpene alcohol (C25) ophiobolin F from *Aspergillus clavatus*.[198] Even though the latter is derived by cyclisation of farnesyl geranyl pyrophosphate (C25, FGPP), the reaction cyclisation mechanism is expected to be comparable to cyclooctat-7-en-3-ol.

The newly assigned stereochemistry for cyclooctat-1,7-diene is (3R*, 6R*, 10S*, 11R*, 14S*) is related to cyclooctat-7-en-3-ol. Both products are derived from premature quenching of the carbocation after the formation of intermediate (**2**). Except for R-cembrene A, all novel terpene produced in this study could be deduced to an incomplete cyclisation cascade. Mechanistic consideration of the mutant structures indicated that the carbocation migration was disrupted in the course of the cyclisation reaction. As a result, alternative reaction pathways to stabilize the carbocation were taking place that were significantly different from the wild type mechanism. These findings are comparable to the reports of mutated sesquiterpene synthases. By contrast, the production of the R-cembrene A represents an exception to this rule. Mechanistic considerations indicated an initial re-arrangement of the GGPP substrate that was unique for this product. Hence, the entire cyclisation reaction is distinct and shares no carbocation intermediates with the parental mechanism.

Using a targeted engineering approach of CotB2, it was demonstrated that diterpene synthases can be engineered for the production of novel structurally distinct macrocycles. The engineering approaches of CotB2 predominantly led to single prominent products, which are suitable for large scale biotechnological productions.

Since the crystal structure of CotB2 was presented in the open and thus inactive form, the substrate bound closed complex of CotB2 was created *in silico*. Based on this data, key amino

acids were analysed that are involved in the cyclisation process. It was found that W288 may be involved in facilitating the 1,5 hydride shift by interacting with the isopropyl group in intermediate (**1**). F107 was suggested to be essential for the orientation of GGDP and involved in mediating the 1,3 hydride shift in intermediate (**2**). Furthermore, F149 was proposed to be essential for stabilizing the carbocation at C3 in intermediate (**3**). However, these statements have to be regarded with caution based on inaccuracy of the *in silico* modelling approach. Nevertheless, this approach enables the selection of amino acid targets for protein engineering with sufficient accuracy compared to open complex crystal structure.

A suitable crystal structure harboring a substrate analog is often unavailable and closed complex crystal structures often contain substrate analogs in a catalytically inactive conformation. As a consequence, the *in silico* modelling approach provides a good structural basis to engineer new diterpene macrocycles.

4. Part III: *Escherichia coli* based Production of Cyclooctatin

Peer-Reviewed Publication

This chapter is based on the following article: "Identification, characterization and molecular adaptation of class I redox systems for the production of hydroxylated diterpenoids" Görner, C.; Schrepfer, P.; Redai, V.; Wallrapp, F.; Loll, B.; Eisenreich, W.; Haslbeck, M.; Brück, T. *Microb. Cell Fact.* **2016**, *15*, 86.

Contributions

CG, PS and TB conceived the study, planned and supervised experiments. PS, CG and VR conducted metabolic engineering of *E.coli* and enzyme mutagenesis experiments. MH, CG and PS conducted fermentation and CD measurements. PS, CG and WE conducted NMR data collection and solved structures of fermentation products. PS, FW and BL conducted *in silico* modelling and verified structures. TB, CG and PS finalized the manuscript.

4.1 Introduction

In the previous chapter the diterpene synthase CotB2 from the cyclooctatin gene cluster (Figure 44) of the soil bacterium *Streptomyces melanosporofaciens* was engineered to produce new and non-native diterpene macrocycles. However, the bioactivity of terpenes relies in most cases on the decorations of functional groups. In the case of the diterpene cyclooctatin, two subsequent hydroxylations are needed to obtain biological activity as human lysophospholipase inhibitor.[10]

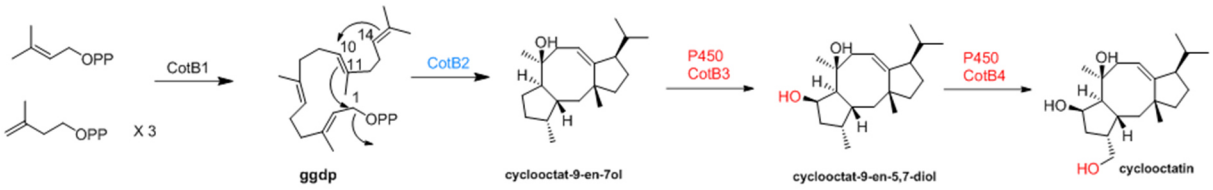


Figure 44. The cyclooctatin biosynthesis in *Streptomyces melanosporofaciens* MI614-43F2. Cyclooctat-9-en-7-ol is the cyclisation product of the diterpene synthase CotB2, which is further functionalized by the P450 monoxygenases CotB3 and CotB4.

The research described in this chapter is dedicated to the production of cyclooctatin in *E.coli*. by reconstitution of the entire biosynthetic gene cluster. The whole cell production of cyclooctatin achieved in this study provided a 43 fold increased yield compared to the native producer. Currently, there is no therapeutic use available for cyclooctatin. However, the results from this research allow the production of cyclooctatin in sufficient quantities to perform a

detailed pharmacological analysis. Furthermore, the genetic engineering approaches developed in this study are suitable for the efficient production of hydroxylated diterpenes from bacterial origin and are therefore a step towards the sustainable production of high value compounds in a bio-refinery.[21]

4.1.1 P450 Monooxygenases

Cytochrome P450 monooxygenases belong to the P450 superfamily, which is one of the most widespread and diverse enzyme systems found in nature.[201] P450 monooxygenases are heme thiolate containing proteins that catalyze the introduction of a single oxygen atom from molecular oxygen (O_2) to an organic substrate, while reducing the other oxygen atom to water.[201] P450 monooxygenases catalyze a variety of region and stereospecific oxidations of hydrocarbons. Reactions encompass oxidation of sp^3 hybridized C atoms, epoxidation of C=C double bonds, aromatic hydroxylation, N-oxidation, deamination, dehalogenation and dealkylations.[201,202,203] Substrates utilized by P450 monooxygenases are highly diverse.[204] Examples are terpenes, fatty acids, steroids, prostaglandins, polyaromatic and heteroaromatic compounds as well as organic solvents.[204] Furthermore, P450 monooxygenases are involved in drug metabolism and detoxification reactions.[201]

P450 monooxygenases utilize O_2 and two electrons, which are derived from the cofactor NAD(P)H and provided by redox partner proteins.[205] The overall equation for the reaction is:[205]

$A + O_2 + NAD(P)H/H^+ + 2e^- \rightarrow AO + H_2O + NAD(P)^+$

4.1.2 P450 Monooxyagenase Reaction Mechanism

The active center of P450 monooxygenases is characterized by a heme prosthetic group (iron protoporphyrin IX).[205] This heme group is identical to the heme moiety found in hemoglobin or other heme containing oxidoreductases such as the peroxidase enzyme family, which all harbor a central iron atom that is ligated to six ligands.[205] The planar porphyrin ring comprises four nitrogen atoms as ligands. A fifth proximal axial ligand consists of a deprotonated cysteine residue.[205] The catalytic cycle of a P450 monooxygenase is illustrated in figure 45.

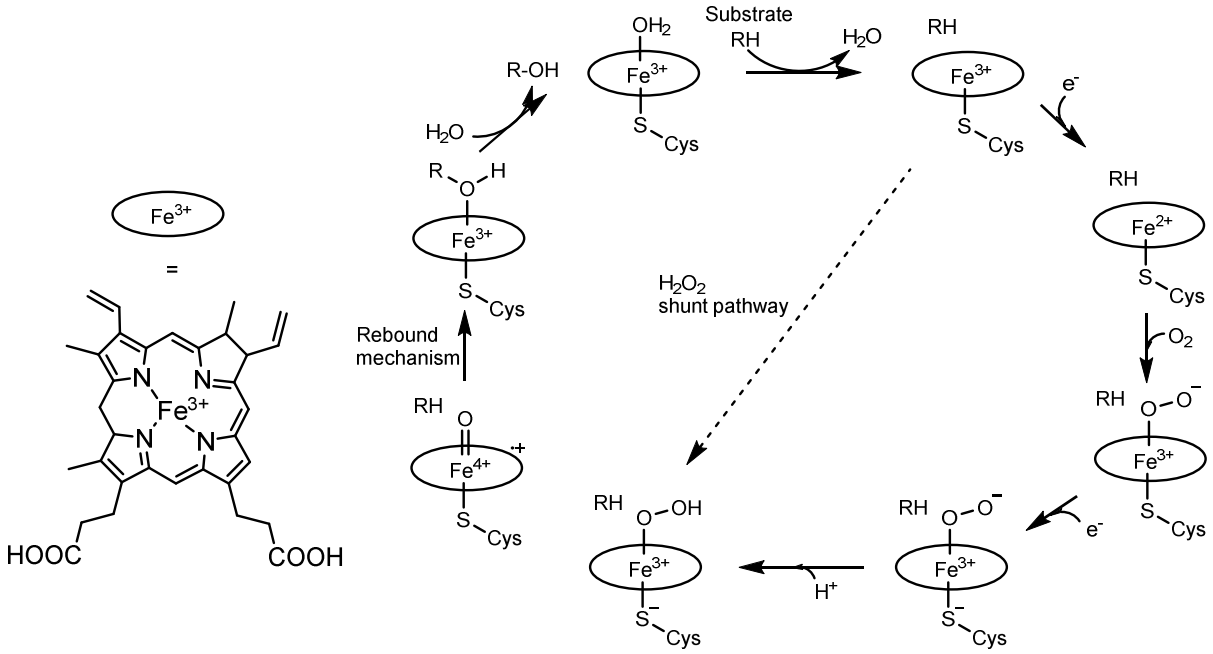


Figure 45. P450 monooxygenase catalytic cycle of oxygen activation, substrate hydroxylation and the shunt pathway. Adapted from [205,206]

In the resting state with no substrate bound, the sixth distal coordination site is occupied by an easily exchangeable water molecule.[205] In this state the iron cation is present in the ferric Fe^{3+} low-spin (S = ½) configuration.[205] After binding of the substrate the water ligand is displaced. The iron cation changes to high-spin (S = 5/2) configuration thereby sinking deeper into the heme group (out-of plane structure).[205] As a consequence, the redox potential of the iron cation is increased, which facilitates the electron transfer to the heme center and reduces the ferric Fe^{3+} state to the ferrous Fe^{2+} state.[205] The substrate induced increase in redox potential resembles an important control measurement to avoid uncontrolled consumption of the NAD(P)H/H^+ cofactor.[205] The reduced ferrous Fe^{2+} state has a high affinity to molecular oxygen. Binding of oxygen produces the relatively stable nucleophilic ferrisuperoxo anion complex (Fe^{3+}-O_2^-).[205] Transfer of the second electron generates a supernucleophilic bi-negatively charged ferriperoxo intermediate ($^-$Cys-Fe^{3+}-O_2^-).[205] The negative charges are located at the dioxygen and are delocalized over the cysteine ligand.[205] Next, a proton is added to the distral oxygen atom to produce a ferrihydroperoxo intermediate ($^-$Cys-Fe^{3+}-O_2H).[205] The distral oxygen is protonated a second time, which leads to the cleavage of the heterolytic O-O bond that produces H_2O and a porphyrin π radical iron (IV) oxo intermediate known as compound I (Por $^{\cdot+}Fe^4$=O).[205] If a C-H bond is oxidized, the monooxygenation reaction proceeds via a hydrogen atom abstraction/oxygen rebound mechanism.[205] In this case the Fe^4=O species of compound I abstracts a H-atom from the substrate (A-H) yielding an alkyl

radical ($A^{\cdot}$) and an iron (IV) bond hydroxyl radical ($PorFe^{4}$-OH).[205] The allyl radical ($A^{\cdot}$) subsequently recombines with the iron (IV) bond hydroxyl radical by the oxygen rebound reaction yielding the hydroxylated substrate A-OH and the heme center in the ferric Fe^{3+} state. The P450 in the Fe^{3+} state can undergo another catalytic cycle. [205,206]

An alternative reaction to yield compound I is the shunt pathway (Figure 45).[205,206] In this pathway the ferric P450 Fe^{3+} reacts with H_2O_2 to form the $^{-}Cys\text{-}Fe^{3+}\text{-}O_2H$ species, which dissociates to compound I after protonation.[205,206] Some P450 monooxygenases are able to utilize H_2O_2.[206] However, in these cases the shunt reaction is inefficient and the catalytic activities of the P450 monooxygenases is lower compared to the $NAD(P)H/H^+$ dependent reaction pathway.[206] In contrast to the H_2O_2 dependent peroxidases, peroxygenases and catalases, P450 monooxygenases are missing acid-base amino acids catalysts such as histidine, aspartic acid and glutamic acid at the distal side of the heme group.[206]

4.1.3 P450 Monooxyagenase Redox Systems

P450 monooxygenases utilize two electrons from the cofactor $NAD(P)H/H^+$. P450 monooxygenases are classified according to the redox system variants that mediate the electron transfer.[207] To date 10 electron transfer chains are known. As the number of characterized P450 systems increases, further redox systems may be discovered.[207] The three most prominent systems are shown in figure 46.[207]

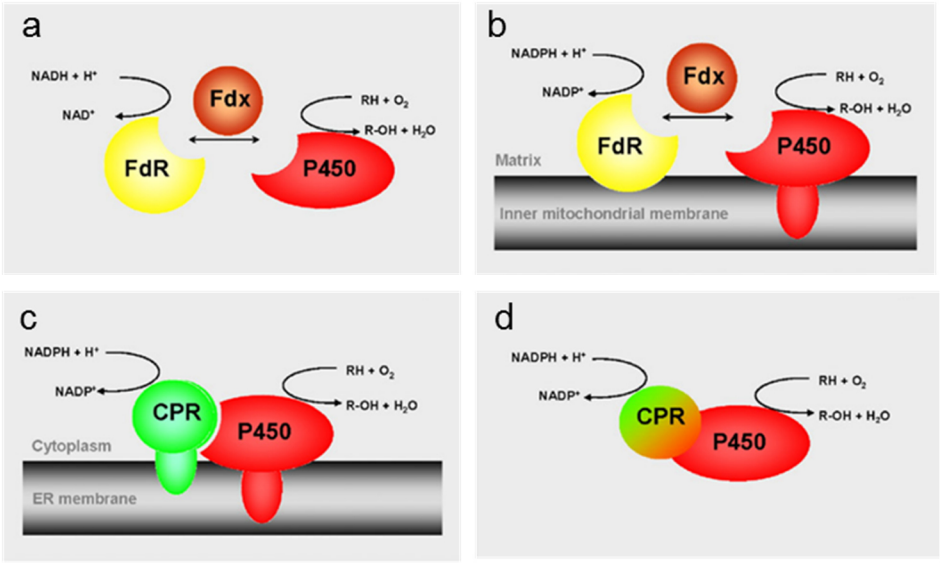


Figure 46.[207] Schematic organization of cytochrome P450 systems. (**a**) Bacterial class I system and (**b**) eukaryotic mitochondrial class I systems.(**c**) Class II microsomal system. (**d**) Bacterial class III systems.

Class I P450 monooxygenases encompass most bacterial P450 systems (Figure 46a) and mitochondrial P450 systems in eukaryotes (Figure 46b).[207] Both systems consist of three different proteins, a FAD containing reductase harboring the NAD(P)H/H^+ cofactor, a ferredoxin containing an [F_2S_2] iron sulfur cluster and the P450 monoxygenase.[207] In the reductase two electrons from NAD(P)H/H^+ are transferred to the FAD cofactor.[208] Next, the ferredoxin binds to the reductase and a single electron is transferred to the iron sulfur cluster of the ferredoxin.[204,208] The ferredoxin shuttles between the reductase and the P450 monooxygenase providing electrons for the oxygenation reaction.[204,208] In bacterial hosts all enzymes of the class I P450 monooxygenase system are soluble.[207] In contrast, in the eukaryotic mitochondria the reductase and P450 monooxygenase are membrane bound and therefore insoluble.[207] Bacterial class I systems are used in the catabolism of carbon sources as well as the production of secondary metabolites such as antibiotics and antifungals.[207] Mammalian mitochondrial class I systems are essential for the biosynthesis of cholesterol derived hormones, vitamin D and bile acids.[207]

Class II P450 monooxygenases are eukaryotic two protein systems, which are insoluble and membrane bound (Figure 46c).[207] Electrons are delivered to the P450 monooxygenase by a P450 reductase, which harbors a NADPH/H^+ binding side as well as the prosthetic groups FAD and FMN.[207] Two electrons are transferred from the NADPH/H^+ over the FAD to the FMN cofactor.[207] Both enzymes are usually located in the endoplasmic reticulum (ER). In plants class II systems are involved in the biosynthesis of secondary metabolites, while fungal systems are essential for the synthesis of membrane sterols and mycotoxins.[207] Mammalian class II systems are crucial for the metabolism of endogenous compounds such as fatty acids, steroids, prostaglandins and exogenous compounds such as therapeutic drugs or toxins.[207]

Class III P450 monooxygenases are monomeric soluble fusion constructs encompassing the P450 reductase and P450 monooxygenase enzyme (Figure 46c).[207] The P450 reductase is N-terminally fused to the oxidase unit and contains a FAD and FMN cofactor analogous to the class II systems.[207] These kinds of P450 enzymes have been discovered in various prokaryotes and lower eukaryotes.[207] The most intensively studied bacterial representative is CYP102A1 (P450BM3) of the soil bacterium *Bacillus megaterium*.[207] P450BM3 is a fatty acid hydroxylase with a molecular weight of 119 kDa.[207] Due to the covalent fusion of the reductase to the P450 monooxygenase, the electron transfer is significantly faster compared with other P450 enzyme systems.[207] Consequently, to-date P450BM3 shows the highest monooxygenase activity of any

P450 systems investigated.[207] P450BM3 has been heavily engineered to hydroxylate alternative substrates.[207] An eukaryotic counterpart of P450BM3 is P450foxy derived from the ascomycete fungus *Fusarium oxysproum*.[207] It shows high primary sequence similarity to P450BM3 and is involved in the terminal oxidation of fatty acids.[207]

4.1.4 P450cam from *Pseudomonas putida* as Prototype of Bacterial P450 Systems

The P450cam monooxygenase from *Pseudomonas putida* is the best characterized P450 member of the cytochrome P450 family.[209] It belongs to the class I bacterial derived P450 systems and stereoselectively hydroxylates (+)-camphor to 5-exo-hydroxycamphor.[12,209] P450cam and its associated redox partners the reductase PdR and the ferredoxin Pdx are used to understand the basis of P450 catalysis.[12]

In 1987 the crystal structure of P450cam was solved. This milestone allowed the first structural characterization of a cytochrome P450.[209] Today more than twenty P450 enzymes of various redox systems have been crystalized.[209] Despite a low primary amino acid sequence similarity of less than 20%, P450 crystal structures share highly conserved elements.[210,211] Figure 47 displays the secondary structure elements of a P450 monooxygenase in a topology diagram.[211]

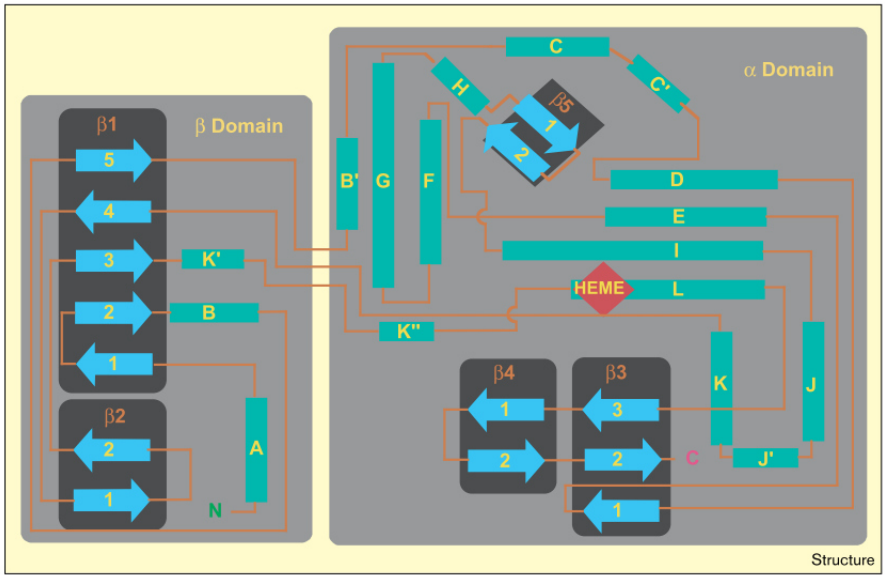


Figure 47.[211] Topology drawing illustrating the secondary structure elements in P450 enzymes. The elements are divided into a α and β domain. Helices are depicted as green rectangles and β sheets are shown as blue arrows. Orange lines represent random coil structure elements connecting secondary structure elements.

P450 enzymes comprise a α and β domain.[211] The α-domain contains the α-helices B' to L and the β-sheet 3-5.[211] The β-domain harbors the β-sheets 1 and 2 as well as the α-helices A, B,

K'.[211] The highest structural homology is found at the catalytic center harboring the heme group.[211] The conserved P450 structural core consists of the four helices bundle D, E, I, L, the helices J, K, the β-sheets β1, β2 and a highly conserved heme stabilizing coil termed the 'meander', which is located on the proximal face of the protein.[211] The β-sheets β1 and β2 are involved in the formation of the hydrophobic substrate access channel.[211] The heme group is located between the distal I- and proximal L- helices. Prior to the L-helix, the highly conserved Phe-X-X-Gly-X-Arg-X-Cys-X-Gly sequence contains the cysteine, which forms the fifth proximal ligand for the iron center.[211] Additionally, the helix-K contains the conserved Glu-X-X-Arg motif, which is essential to stabilize the core heme binding structure.[211] The conserved sequence Ala/Gly-Gly-X-Asp/Glu-Thr-Thr/Ser in the I-helix contains the essential amino acids Asp/Glu and Thr, which are involved in the protonation and subsequent water cleavage of the ferriperoxo intermediate ($^{-}$Cys-Fe^{3+}-O$_2^{-}$) to yield compound I.[205,207]

4.1.5 Pseudomonas Putida Ferredoxin Pdx

Ferredoxins harboring a [2Fe-2S] cluster are found in bacteria, plants and higher animals.[212] Two classes of [2Fe-2S] ferredoxins are reported, the photosynthetic plant-type and oxidase coupled vertebrate-type ferredoxins.[212] While the plant-type ferredoxin is involved in electron transfer from the photosystem I to the ferredoxin-NADP$^+$ reductase, the vertebrate-type transfers electrons from the NAD(P)H/H$^+$ dependent reductases to cytochrome P450 enzymes.[212] The latter is structural homologs to bacterial [2Fe-2S] ferredoxins including the ferredoxin Pdx from *Pseudomonas putida* involved in the P450cam redox system.[212] The crystal structures of the reduced and oxidized Pdx has been solved (Figure 48).[213]

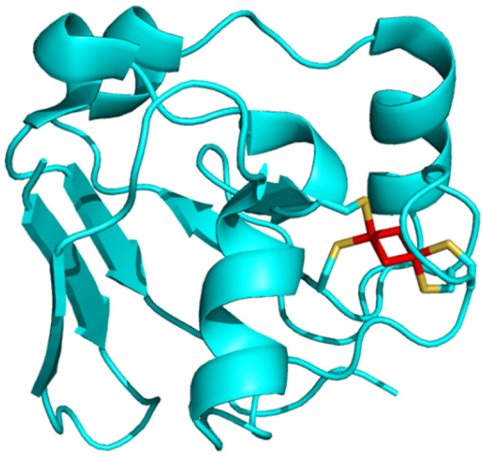


Figure 48. The structure of the reduced *Pseudomonas putida* ferredoxin Pdx depicted in cartoon representation (PDB entry 1XLQ) and colored in cyan.[213] The iron sulfur cluster is colored in red. Cysteine side chain residues assembling the iron sulfur cluster are colored in yellow.

Pdx was found to have a similar fold compared to the *E.coli* ferredoxin (Fdx) and the bovine mitochondrial adrenodoxin (Adx)[212] Pdx is a small protein and consists of 107 amino acids, which forms a hydrophobic core structure from a five-stranded β-sheet.[212] The β-sheet structure is flanked by three small α-helices and a protruding C-terminal loop (Figure 48).[212]The iron sulfur cluster consists of the residues Cys39, Cys45, Cys48, and Cys86 and is located on the surface of the protein.[212] The [2Fe-2S] cluster contains two high spin iron ions, which are present in the Fe^{3+} Fe^{3+} state when oxidized and in the Fe^{3+} Fe^{2+} state when reduced.[212] Pdx was shown to adapt a minor conformational change when reduced.[212]

4.1.6 Binding and Electron Transfer from the *Pseudomonas putida* Reductase PdR to the Ferredoxin Pdx

Binding between Pdx and PdR as well as P450cam was reported to be rather weak.[12] The inner view showing essential amino acids between the putida reductase PdR and ferredoxin Pdx binding complex is illustrated on Figure 49.[12]

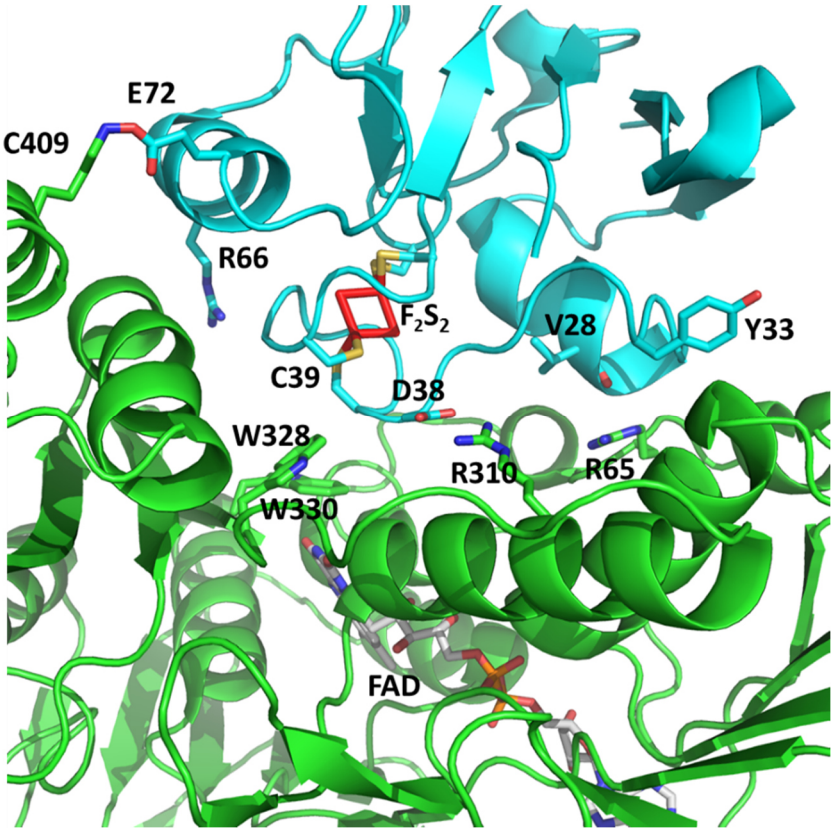


Figure 49. Inner view of the structure showing the cross linked complex (PDB entry 3LB8 from the *Pseudomonas putida* reductase PdR (green) and ferredoxin Pdx (cyan).[12] The iron sulfur cluster and FAD cofactor are illustrated. Essential amino acids at the Pdr–Pdx interface are highlighted. Note that K409 is covalently bound to E72.[12]

Binding of Pdx to PdR is mostly facilitated by two ionic interactions between $Arg310^{PdR}$ / $Asp38^{Pdx}$ and $Lys409^{PdR}$ / $Glu72^{Pdx}$ as well as the hydrogen bond between $Arg65^{PdR}$ and the amino acid backbone of $V28^{Pdx}$.[12] At the Pdr-Pdx interface the two residues $Trp328^{PdR}$ and $Trp330^{PdR}$ form a direct link between the Flavin (PdR) and iron-sulfur (Pdx) cofactors.[12] In contrast, the bulky amino acids Tyr33, Arg66 and Trp106 (not shown in the PdR-Pdx complex) in the ferredoxin Pdx prevent tight binding and facilitate dissociation upon reduction.[12] The electron transfer between the FAD cofactor in PdR and the [2Fe-2S] in Pdx is assumed to include $Asp38^{Pdx}$ and $Cys\ 39^{Pdx}$.[12]

4.1.7 Binding and Electron Transfer from Pdx to P450cam in *Pseudomonas putida*

In the resting state P450cam adopts the open conformation. In this state water is bound to the iron center, which is in the low-spin state.[12,214,215] Substrate binding facilitates the structural rearrangement to the closed conformation, which is characterized by the high-spin state of the iron center. Alternatively, the presence of reducing agents forces P450cam to adopt a closed conformation in the absence of a substrate.[215] While a reducing agent can transfer the first electron to the iron center thus reducing the ferric Fe^{3+} to the ferrous Fe^{2+}, the second electron transfer for the formation of complex I is exclusively facilitated by the Pdx ferredoxin.[12] Pdx binds on the proximal site of P450cam, which is the opposite site of the substrate channel.[12]

Interestingly, when reduced Pdx binds to the substrate-bond closed conformation of P450cam resulting in the P450cam·Pdx complex, P450cam in this complex adopts the open conformation.[214] This conformational switch is believed to be essential to establish the second electron transfer.[214] Additionally, formation of the open conformation enables catalytic water to enter the active site, which establishes a proton relay network needed for the cleavage of the O-O bond and thus the formation of compound I.[214] Structural superpositioning of P450cam from the substrate bond P450cam structure (closed conformation) and the P450cam·Pdx complex in the open conformation, reveals a rearrangement of the helix C leaning towards Pdx followed by shifts in the helices I, B', F and G as illustrated on figure 50a.[214] An inner view highlighting essential amino acids for the binding and electron transport is shown in figure 50b.[214]

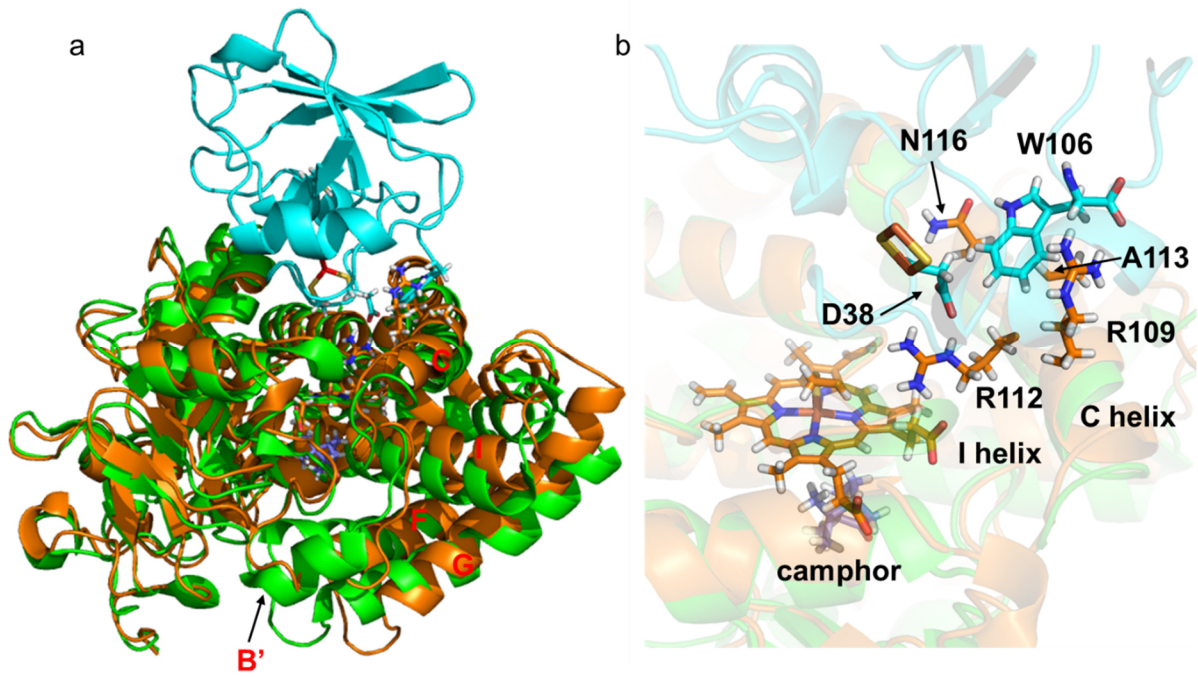


Figure 50. (**a**) Superposition of the P450cam·Pdx complex (P450cam: orange; Pdx: cyan) (PDB entry 4JX1) forming the open conformation with the closed conformation of single P450cam (substrate bound) (green) (PDB entry 2ZWT).[214] Large alterations are visible at the F/G loop connecting the F and G helices, the B′ helix, which is disordered in the open configuration. Furthermore, movements in the I and C helix are observed. The C helix in the P450cam·Pdx complex is shifted towards Pdx. (**b**) Inner view of (**a**) illustrating essential amino acids for the binding and electron transport between P450cam and Pdx.

Key amino acid interactions between Pdx and P450cam are the salt bridge between $Asp38^{Pdx}$ and $Arg112^{p450cam}$, the hydrophobic interactions between $Trp106^{Pdx}$ and $Ala113^{p450cam}$ as well as the H-bonds between $Trp106^{Pdx}$ and $Asn116^{p450cam}$.[214] Furthermore, $Arg109^{p450cam}$ interacts with the C-terminal carboxyl group of $Trp106^{Pdx}$.[214] An electron transfer route between Pdx and P450cam is proposed from the iron sulfur cluster to $Asp38^{Pdx}$ over $Arg112^{p450cam}$ to the heme group.[214] Interestingly, the first electron transfer was reported to be independent of the C-terminal $Trp106^{Pdx}$, while for the second transfer $Trp106^{Pdx}$ is essential.[12,214] Consequently, $Trp106^{Pdx}$ plays a crucial role for the close-to-open conformational change upon binding of PdR and mediates the transfer of the second electron.[214]

The high diversity and promiscuity of P450s is believed to be caused by variations in the structure of the substrate binding channel, which encompasses the helices I, B′, F, G and the FG loop.[211] Variations in the amino acid side chains in these structural elements result in P450s with distinct substrate bindings specificities.[211] Additionally, conformational changes due to substrate binding as well as binding of the corresponding redox partners play an important role in P450 substrate access and recognition.[211,214] In this respect, the well characterized P450cam system can serve as model for P450s.

4.2 Scope of This Work

Cytochrome P450 enzymes are essential for oxidation of carbon skeletons in the biocatalysis of highly functionalized and complex secondary metabolites.[210] Consequently, for the biotechnological production of secondary metabolites the functional reconstitution of P450 enzyme systems is crucial.[201,204,210] Secondary metabolites from the bacteria genus *Streptomyces* comprise a series of terpene and polyketide based bioactives with antifungal, antimicrobial and anti-tumor activates.[216,217,218,219] Cultivation of these *Streptomyces* strains often results in low yields and a high product mixture that resembles a challenge for the biotechnological production.[220] An alternative to the native post production, is the heterologous reconstitution in the well characterized bacteria *E.coli*.[220,221] Recent successes in reconstitution of upstream pathways have been achieved for the production of terpene and polyketide based macrocycles.[132,222,223,224,225,226] However, the reconstitution of the subsequent downstream pathway involving P450 system can be challenging. In contrast to plant derived P450 system, the transfer of bacterial P450 system into *E.coli* is much simpler, since all participating enzymes are usually soluble.[158] Nevertheless, the bottleneck in this approach is to find the corresponding P450 dependent redox systems consisting of the ferredoxin and reductase, which often remains elusive.[12] Alternatively, the reconstitution of P450 oxygenases can be done by using structurally related non-native redox systems such as the flavodoxin and flavodoxin reductase from *E.coli*, the putidaredoxin and putidaredoxin reductase from *Pseudomonas putida*, the bovine adrenodoxin reductase (AdR) and bovine adrenodoxin (Adx) or the *spinach* ferredoxin (spFdx) and spinach ferredoxin reductase (spFdR).[227,228,229,230,231] Nevertheless, the catalytic efficiency of the cytochrome P450 enzymes is case dependent and often lower compared to the native redox systems. Although one report describes an increased catalytic efficiency by using a non-native redox system compared to the native counterpart.[232] Additionally, the catalytic efficiency of P450 oxygenases can be increased by engineering the electron transfer between the P450 hydroxylase and ferredoxin using amino acid mutagenesis.[233,234]

In this study the cyclooctatin gene cluster is reconstituted in *E.coli*. The gene cluster contains two class I P450 enzymes CotB3 and CotB4, which are missing in the corresponding redox system. Consequently, the challenge is to find a suitable non-native redox system which sufficiently mediates the electron transfer to the P450 enzymes, which only share 51% primary

sequence homology. Additionally, a further task is to reconstitute the upstream metabolic pathway for the diterpene production as well as the downstream pathway for the subsequent P450 mediated hydroxylations in the same *E.coli* host. A series of researchers have reported the successful production of hydroxylated mono- and sequiterpenes in *E.coli*.[9,204,225] In contrast only few publications report the hydroxylation of diterpenes.[146,232,235] The most promising example to date is the hydroxylation of the diterpene taxadiene to taxadiene-5α-ol by a plant derived class II P450 system that was reconstituted by a fusion construct between corresponding reductase and P450 monooxygenase.[146] Unfortunately, in this example the production yield of the precursor taxadiene dropped significantly in presence of the P450 system.[146] Another milestone is the consecutive two times hydroxylation of the polyketide Erythromycin A.[222] Interestingly, the *E.coli* native class I reductase and ferredoxin system was sufficient to achieve activation of these two P450 enzymes.[222] Nevertheless, this example proves the feasibility to reconstitute multiple P450 enzymes at the same time.

The successful production of cyclooctatin is the first account for a complete, heterologous biosynthesis cascade of a multi-functionalized diterpenoid in *E.coli*. Moreover, it is an important step towards the large scale production of Streptomyces derived secondary metabolites. Methods and approaches applied in this study are generally applicable for the biotechnological production of highly functionalized bioactives .

4.3 Results and Discussion

4.3.1 Identification of Redox Partners

The bioinformatics to identify and model the reductase (AfR) and ferredoxin (Afx) from *Streptomyces afghaniensis* was executed by Patrick Schrepfer and is part of the publication: Görner, C.; Schrepfer, P.; Redai, V.; Wallrapp, F.; Loll, B.; Eisenreich, W.; Haslbeck, M.; Brück, T. *Microb. Cell Fact.* **2016**, *15*, 86.

To identify suitable redox partners for CotB3/B4 a multistep computational strategy was adopted. The idea was to find a suitable class I redox system from taxonomically related *Streptomyces* strain. Therefore, the *P. putida* derived reductase (PdR) and ferredoxin (Pdx) were used as templates for a homology-bases search covering all available *Streptomyces* genomes.

Unfortunately, the genomic data of the native cyclooctatin producer *S. melanosporofaciens* was not available and could not be taken into account. The homology search indicated one putative PdR and several hypothetical Pdx homologs with up to 45% sequence similarity, which all clustered in the recently shotgun-sequenced genome of *Streptomyces afghaniensis*. In the next step, the sequences were subject to secondary structure prediction using the HMM (Hidden Markov Model)-HMM comparison by the HHpred server tool.[174] Protein models were created by using the best hits including the redox system of *P. putida*, the mitochondrial adrenodoxin redox system and the terpredoxin system from *Pseudomonas sp*. The predicted protein models were structurally fine-tuned by several energy minimization and MD-simulation steps using the Yasara bioinformatics Suite. To restrict the selection, the optimized 3D models were aligned to the *P. putida* class I redox equivalents, which lead to the identification of a NADH-dependent ferredoxin reductase (afghanoredoxin reductase, AfR) and its corresponding 2Fe-2S ferredoxin (afghanoredoxin, Afx). Interestingly, the primary amino sequences of the new AfR (47%) and Afx (38%) proteins differed significantly from the *P. putida* PdR/Pdx model, while the modelled protein folding was found to be very similar. Figure 51 shows the structural alignment of AfR/PdR and Afx/Pdx.

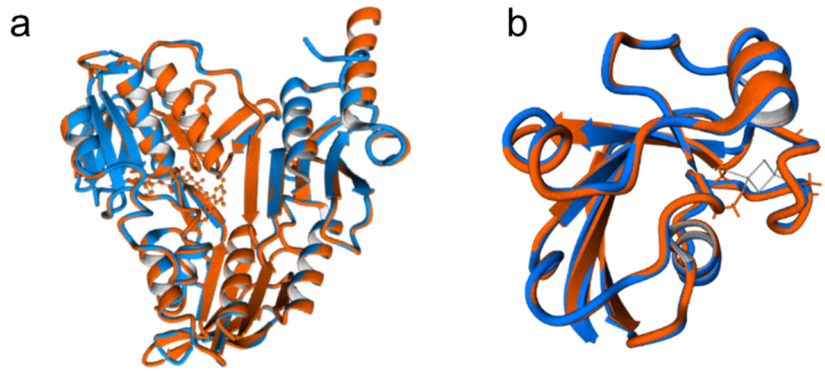


Figure 51. Superposition of AfR (orange) with PdR (blue) (**a**) and Afx (orange) with Pdx (blue) (**b**).

Despite the low primary sequence similarities both alignments have a root mean square deviation (r.m.s.d) of 0.1 Å over 404 amino acids for the reductase and 106 aligned amino acid residues for the ferredoxin.

To test the redox system AfR/Afx, bicystronic operons comprising AfR/Afx and P450 CotB3/CotB4 were synthesized and introduced into the plasmid pACYC-Duet-1. The plasmid was chosen, since it is compatible with the previous used *E.coli in vivo* screening system. Furthermore, the plasmid displays a low copy number, which translates into a decreased expression level that was shown to be beneficial for the production of hydroxylated terpenoids.[146] The *E.coli* strain Bl21 (DE3) used for the *in vivo* screening harbored the plasmids pColaDuet-1(DXS/DXP), pCDFDuet-1(IspD/IspF and IDI) and pETDuet-1 (CrtE/CotB2). The newly created plasmid pACYC-Duet-1(AfR/Afx, CotB3/CotB4) was introduced into this strain and cultivated in shake flasks (400 mL) for 5 days. Terpenes were extracted from the cell pellet as well as the growth media and analyzed by GC-MS/FID. It was anticipated that adding additional hydroxyl groups to cyclooctat-9-en-7-ol significantly increases the boiling point thus making normal temperature GC analysis impossible. Therefore, all samples were treated with BSTFA to transform all hydroxyl groups into volatile trimethyl-silyl ethers. GC-MS/FID analysis indicated the production of three different terpene species corresponding to cyclooctatin (1.6 mg/L) and its biosynthetic intermediates cyclooctat-9-en-7-ol (1.1 mg/L) and cyclooctat-9-en-5,7-diol (0.8 mg/L), which showed the corresponding molecular mass [M^+] of 362 Da, 466 Da and 538 Da (Appendix Figure 77-79). For better illustration, Figure 52 shows the GC-FID spectrum of the terpenes extracted from a 48 hours bioreactor cultivation.

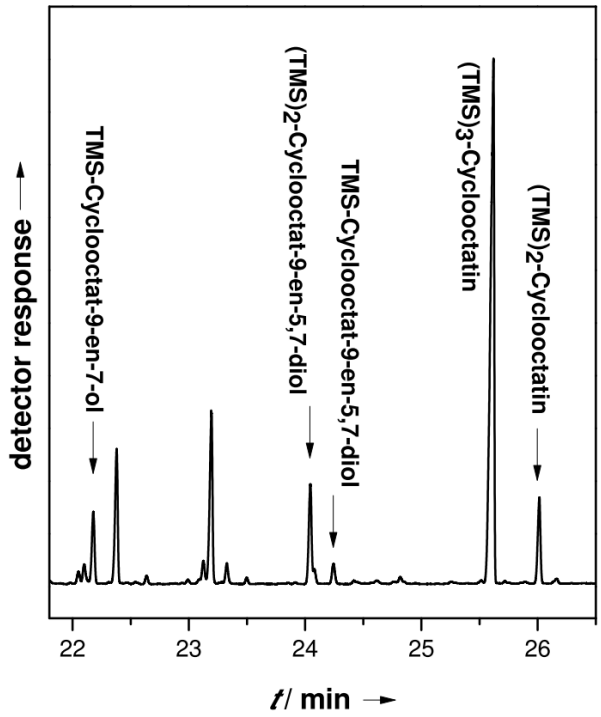


Figure 52. GC-FID chromatogram of a silylated cellular extract from the cyclooctatin producing *E.coli* strain in a bioreactor harvested after 48 hours. (Yields: cyclooctatin 6.9 mg/L, cycooctat-9-en-5,7-diol 1.15 mg/L, cyclooctat-9-en-7-ol 0.64 mg/L)

4.3.2 Determination of the Absolute Configuration of C-5 in Cyclooctat-9-en-5,7-diol

Interestingly, no product isomers were detected, which implies that the hydroxylase activities of CotB3 and CotB4 were absolutely stereoselective. To provide sufficient material for a detailed 1D-/2D NMR based structure elucidation (Appendix Figure 49-55), the cultivations were repeated several times. The ^{1}H/ ^{13}C chemical shifts and NOE correlations were in good agreement with literature values.[10] Furthermore, the absolute stereochemistry of the hydroxyl group at C-5 was analyzed by Mosher`s NMR methodology.[236] Treatment of cyclooctat-9-en-5,7-diol with the (R) and (S)-MTPA chloride provided the (S)- and (R)-MTPA product. The esterification was specific for the secondary hydroxyl group at C-5, while leaving the ternary hydroxyl group at C-7 unaffected. Calculation of the Δδ S - R values, shown in figure 53 (Appendix Figure 56-65, Appendix Table 3, 4) clearly define the absolute configuration of C-5 as R.

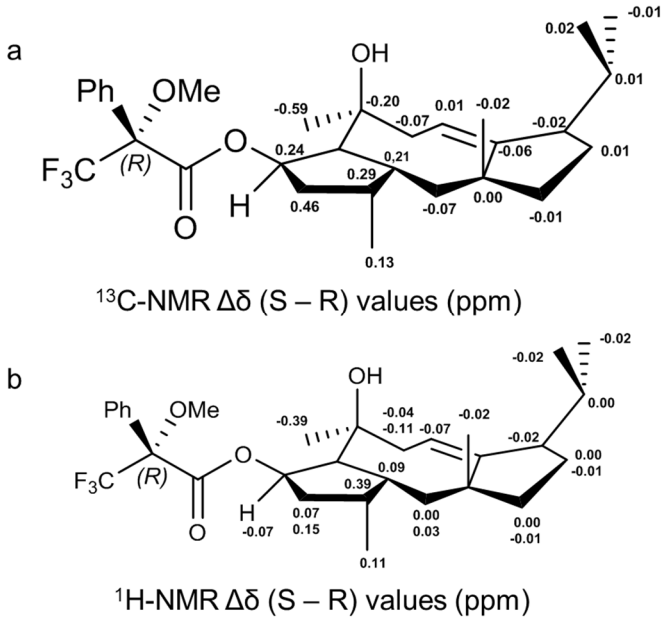


Figure 53. The calculation of the Δδ S - R values of the (S)- and (R)-MTPA products of cyclooctat-9-en-5,7-diol deduce the absolute configuration of the hydroxyl group at C-5 as R.

The absolute configuration of cyclooctatin was previously suggested as 2R-, 3R-, 5R-, 6S-, 7S, 11R, 14R[198] and is in agreement with the absolute stereochemistry determined for C-5 in this study.

4.3.3 Fermentative Production of Cyclooctatin

Scaling up the cyclooctatin production to 1L shake flask cultures did not increase the yields of the tri-hydroxylated product. Instead, it was observed that biosynthetic intermediates accumulated, while the cyclooctatin yields decreased. Consequently, cultivations were proceeded in a 5 L bioreactor system with controlled oxygen supply. Cultivation was done in LB medium for 4 days at 25°C with an O_2 saturation of 80%. The fermentation was done as a batch process with initial 40 g/L glycerol as carbon source. The amount and distribution of cyclooctatin and its metabolic precursors was analyzed every 24 hours in the *E.coli* cells as well as in the cell free fermentation medium. The results are illustrated in figure 54.

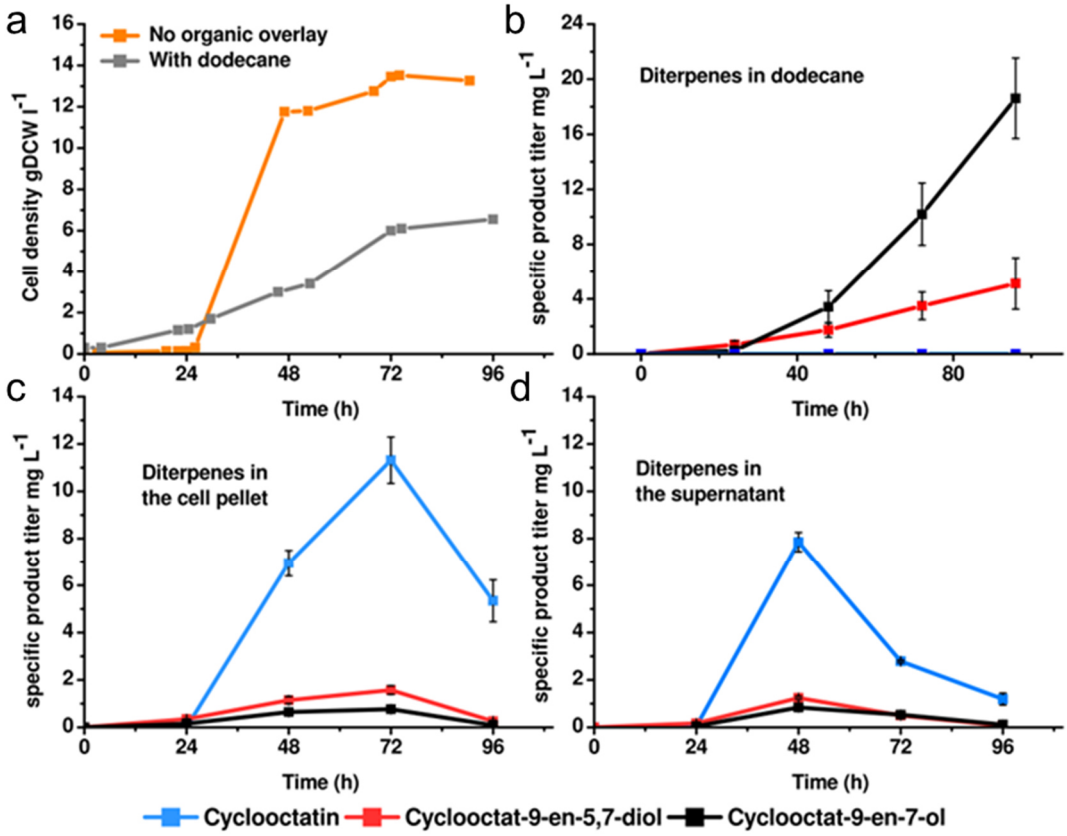


Figure 54. (**a**) Cell density during fermentation in the presence and absence of a dodecane organic phase overlay. (**b**) Production of cyclooctatin and its biosynthetic intermediates in the dodecane organic phase is shown. (**c,d**) Diterpene production during fermentation in absence of dodecane.

The fermentation procedure was based on Boghigian et al.[237] Cultivation started with an inoculation of 0.1 (OD 600) at 25°C in the presence of 1 mM IPTG.[237] Growth curves and time dependent product yields are illustrated in figure 54. In the first 24 hours the cell growth showed an elongated lag phase, which is caused by the low temperature combined with the induced protein expression. In the next 24 hours the culture displayed exponential growth, which turned into stationary growth conditions with only minor growth rates. At the end of the fermentation (72 hours), 13.7 g/L dry cell weight biomass was produced. In contrast to the shake flask experiments, cyclooctatin was the major product. The maximal yield was determined after 48 hours with 15 mg/L, which is significantly higher compared to the production in the native *S. melanosporofaciens* (0.35mg/L).[238] The highest content of cyclooctatin was 11 mg/L found after 72 hours in the *E.coli* cell pellet. By contrast, the highest concentration of cyclooctatin detected in the fermentation medium was 8 mg/L after 48 hours. Interestingly, after these concentration peaks the cultivation period was marked by a steady decrease in the cyclooctatin titer. The reasons for this may be oxidative product degradation or product loss by the aeration system. The latter was reported to decrease the titers in the

fermentative production of the diterpene taxadiene and the sequiterpene amorphadiene.[133,146] Unlike these hydrophobic terpenes, the tri-hydroxylated diterpene cyclooctatin is more hydrophilic but still water insoluble. Therefore, the high aeration rates used in this study, which are necessary for efficient P450 enzyme activity may be accountable for the loss. Terpene degradation products were not detected during the fermentation.

To increase the cyclooctatin titers an organic overlay was applied to the fermentation medium. This methodology was reported to increase yields of other oxygenated terpenoids.[133,146] A 1 L cyclooctatin fermentation supplemented with 20% (*v*/*w*) dodecane overlay was carried out. The terpene composition was quantified in the organic phase during the fermentation. The results are shown in figure 54b. Interestingly, the lag phase was significantly reduced and cell growth started rapidly. However, cell growth was significantly slower compared to the 5 L fermentation in absence of the dodecan overlay. After 72 hours, the *E.coli* cells were in stationary phase. At the end of the fermentation 6.5 g/L dried cell biomass was produced. Analysis of the organic phase showed that only the cyclooctatin precursors cycloocat-9-en-7-ol (18.6 mg/L) and cyclooctat-9-en-5,7-diol (5.1 mg/L) were produced. Only trace amounts of cyclooctatin were detected in the organic overlay and in the *E.coli* cells. This finding suggests that the metabolic precursors are rapidly transferred from the cells to the fermentation medium and captured in the organic layer, providing insufficient metabolic flux for the production of cyclooctatin.

Despite the fact that the application of an organic overlay did not increase production titers, the rather simple plasmid based *E.coli* production system was able to provide a 43 fold increased cyclooctatin yield compared to the wild type producer. It can be expected that a plasmid free and genomically engineered *E.coli* can significantly improve the production titers. In particular the application of promotors of various strengths are expected to further increase titers.

4.3.4 The Substrate Promiscuity of CotB3 and CotB4

Hydroxylation of cycooctat-9-en-7-ol by the P450 hydroxylases CotB3 and CotB4 was found to be exclusively stereoselective. The substrate promiscuity depends on the amino acids present in the active center that hold the substrate in close proximity to the iron center. To evaluate the substrate promiscuity of CotB3/4 with respect to alternative diterpene macrocycles, the diterpene synthase CotB2 was exchanged. Diterpene synthases for the production of (-)-casbene, taxa-4,5-11,12-diene as well as mutant CotB2 enzymes to produce R-cembrene A and (1R,3E,7E,11S,12S)-3,7,18-dolabellatriene were evaluated. Therefore, 100 mL shake flask experiments were carried out. After 3 days the growth medium and *E.coli* biomass was extracted and analyzed by GC-MS . In order to analyze potential multiple hydroxylated products all samples were treated with BSTFA prior to analysis. Unexpectedly, a single hydroxylated product was obtained when (-)-casbene was tested. Evaluation of CotB3 and CotB4 as single enzymes revealed that CotB3 was able to hydroxylate about 20% of (-)-casbene in a shake flask culture (400 mL) yielding approximately 1.4 mg/L (Figure 55).

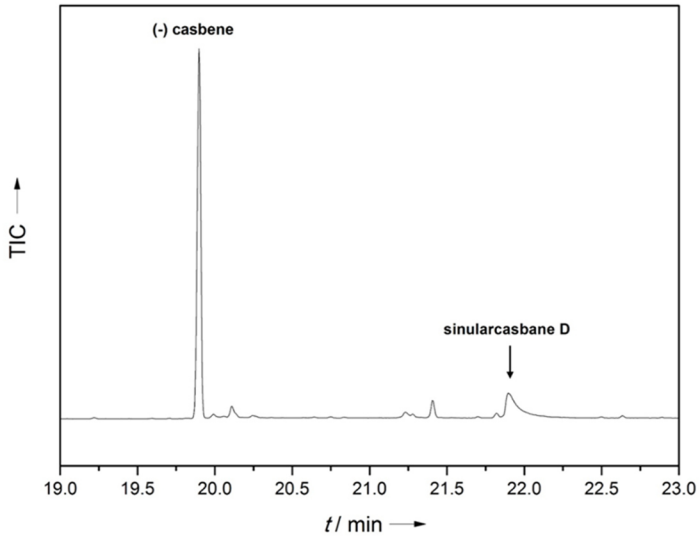


Figure 55. GC-MS spectrum from *E.coli* cell pellet extract of a 72h shake flask culture. The *E.coli* strain produced (-)-casbene and functionally expressed CotB3, which led to the formation of the hydroxylated (-)-casbene - sinularcasbane D.

4.3.5 Structure Determination of Sinularcasbane D

To determine the position and stereochemistry of the hydroxyl group, the product was purified and analyzed by 1D/2D NMR spectroscopy (Appendix Figure 66-74). As reference the metabolic precursor (-)-casbene was also purified and analyzed by NMR spectroscopy. (-)-

casbene suffers from high signal overlap in the COSY and NOESY spectrum. To date a detailed NMR structure elucidation has not been reported. As a consequence, (-)-casbene was identified by fingerprint comparison of the ^{1}H and ^{13}C signals with literature values.[239] The relative configuration could be assigned by NOESY NMR spectroscopy. Therefore, the readily identified proton signals at δ = 0.58 and 1.22 ppm corresponding to either H-1 or H-2 were analyzed in the NOESY spectrum. A clear correlation between both proton signals indicated a *cis*-configuration, which is in agreement with (-)-casbene. Eventually, the absolute configuration was determined by measuring the sign of the optical rotation. A negative rotation value confirmed the structural identify of (-)-casbene. The structurally increased asymmetry of the hydroxylated (-)-casbene allowed a complete structure characterization by NMR spectroscopy (Table 6, Figure 56).

Table 6. NMR data of sinularcasbane D.

#	δ_H (ppm), *J* (Hz)	δ_C (ppm)		#	δ_H (ppm), *J* (Hz)	δ_C (ppm)	
1	0.65 (ddd, *J* = 10.5; 8.7; 1.7, 1H)	30.80	CH	**11**	4.98 (dq, *J* = 9.2; 1.4, 1H)	127.01	CH
2	1.21 (t, *J* = 8.5, 1H)	25.97	CH	**12**	-	140.31	C
3	4.84 (d, *J* = 8.3, 1H)	121.41	CH	**13**	a 2.24 (m, 1H) b 1.81 (m, 1H)	39.70	CH_2
4	-	135.89	C	**14**	a 1.68 (m, 1H) b 1.13 (m, 1H)	24.03	CH_2
5	a 2.19 (m, 1H) b 2.06 (m, 1H)	39.57	CH_2	**15**	-	20.16	C
6	a 2.19 (m, 1H) b 2.05 (m, 1H)	25.09	CH_2	**16**	1.07 (s, 3H)	29.06	CH_3
7	4.87 (ddd, *J* = 6.7; 4.4; 1.2, 1H)	127.09	CH	**17**	0.96 (s, 3H)	15.74	CH_3
8	-	130.95	C	**18**	1.62 (s, 3H)	16.08	CH_3
9	a 2.35 (m, 1H) b 2.08 (m, 1H)	48.15	CH_2	**19**	1.59 (s, 3H)	17.25	CH_3
10	4.52 (ddd, *J* = 10.7; 9.1; 3.2, 1H)	67.00	CH	**20**	1.66 (s, 3H)	18.72	CH_3

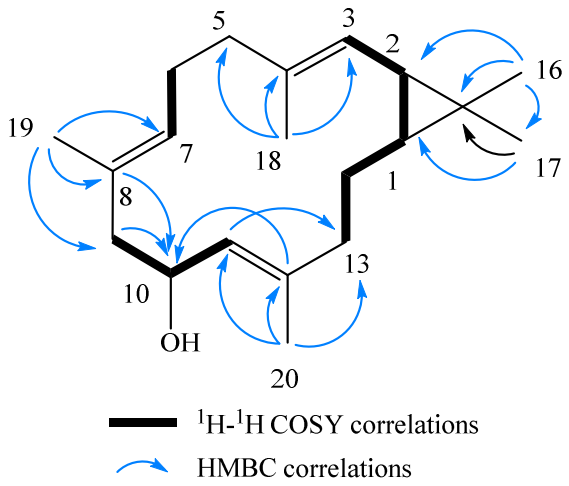


Figure 56. Structure of sinularcasbane D. Key correlations in the COSY and HMBC NMR spectrum are indicated.

^{13}C, Dept 90/135 spectroscopy showed the presence of a new quaternary carbon atom at δ=67.00 ppm with the characteristic chemical shift of an alcohol group. Three proton spin systems were identified by COSY NMR comprising H_2-13/H_2-14, H_2-14/H-1, H-1/H-2, H-2/H-3, H-7/H_2-6 and H_2-9/H-10, H-10/H-11 (Figure 56). By applying the spin systems to the parental (-)-casbene structure it was possible to assign the hydroxyl group to C-10. In the next step the structure was submitted as a search query in the structure database Reaxys.[193] Surprisingly, a compound with identical connectivity was found. The reported compound sinularcasbane D was characterized as a natural product isolate from the soft coral *Sinularia sp*.[14] Comparison of the ^{1}H and ^{13}C signals with literature signals was generally in good agreement, except the ^{13}C signals for C-3, C-11 and C-12, which differed by 0.21, 0.21 and 0.20 ppm to the reference, respectively.[14] As a result it was speculated that the CotB3 hydroxylation product might present a stereoisomer of sinularcasbane D. To determine the stereochemistry of the new alcohol function, Mosher`s NMR methodology was attempted. Unfortunately, treatment of the hydroxylated casbene substance with the (R) and (S)-MTPA chloride led to the decay of the structure. Consequently, NOESY NMR spectroscopy was applied to determine the relative configuration of the hydroxyl group.

Therefore a 3D structure model of sinularcasbane D was created and verified using characteristic NOE correlations demined in the NOESY NMR spectrum (Figure 57).

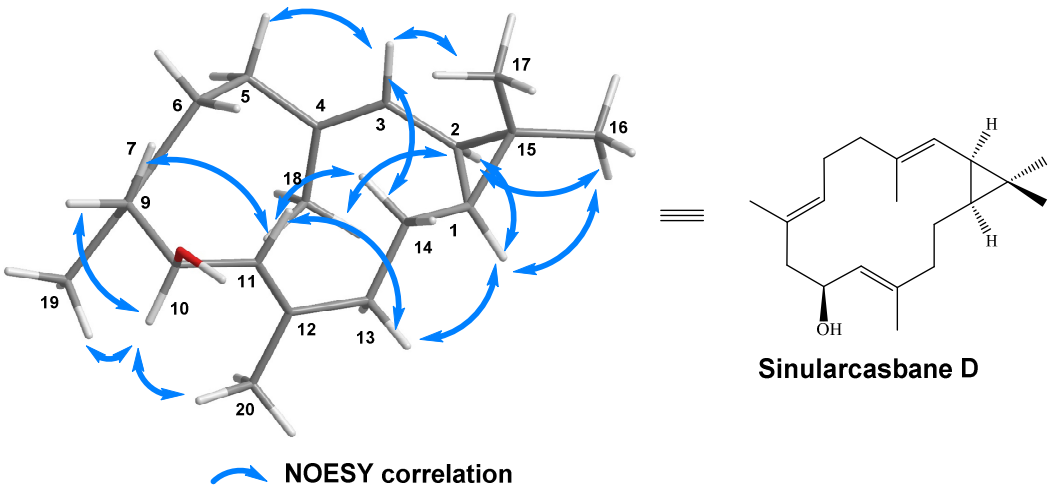


Figure 57. Molecular model (MM2) of sinularcasbane D. Key NOESY correlations are illustrated.

In analogy to the parental compound (-)-casbene, the correlations between H-1 and H-2 were preserved. NOE correlations of H-10 were found between H-10/H-9β, H-10/H_3-20 and H-10/H_3-19. H-10 was assigned to the α-position for the following reasons. In the modelled structure, both methyl groups H_3-19 and H_3-20 are in proximity to H-10α thus making a NOE correlation very likely. Since only H-9α shows a correlation between H-11, the correlation between H-9β and H-10α indicates a *trans*-configuration between H-10α and H-11. These data strongly indicate the presence of H-10 in the α-configuration and the hydroxyl group at C-10 is the β-configuration, which is in agreement with stereochemistry of sinularcasbane D. As the absolute configuration of (-)-casbene was previously determined, the absolute configuration of C-10 was assigned as R.

The hydroxylation of (-)-casbene to produce sinularcasbane D is the first reported biotechnological production of this natural product. This scalable and sustainable production pathway presents an interesting alternative to the isolation from the soft coral *Sinularia sp*. The biological activity of sinularcasbane D has not been extensively characterized. However related natural products have been shown to display moderate cytotoxicity and antimicrobial activity.[14,240]

4.3.6 *In silico* Docking of Sinularcasbane D in CotB3

The bioinformatics described in this chapter were executed by Frank Wallrap and Patrick Schrepfer and is part of the publication: Görner, C.; Schrepfer, P.; Redai, V.; Wallrapp, F.; Loll, B.; Eisenreich, W.; Haslbeck, M.; Brück, T. *Microb. Cell Fact.* **2016**, *15*, 86.

To elucidate the structure function relationship of the P450 enzyme CotB3 and the production of sinularcasbane D, a model structure of CotB3 was created. Therefore, the structure of the *Streptomyces coelicolor* cytochrome P450 (PDB 3EL3, 39% sequence identity) was used as main template and the crystal structure of the human cytochrome P450 (PDB 3LD6, 23% sequence identity) was used as template for the missing loops between the amino acids 194 -209 and 250 -258. In the next step, a structure model of (-)-casbene was created and docked into the CotB3 structure. Figure 58 shows (-)-casbene in the active site of CotB3 adapting the orientation with the lowest determined binding energy. Interestingly, in this conformation (-)-casbene establishes the optimal positioning for the hydroxylation at C-10. It is notable that (-)-casbene adopts a unique skewed orientation, which might explain why this is the only alternative diterpene macrocycle that was subject to hydroxylation.

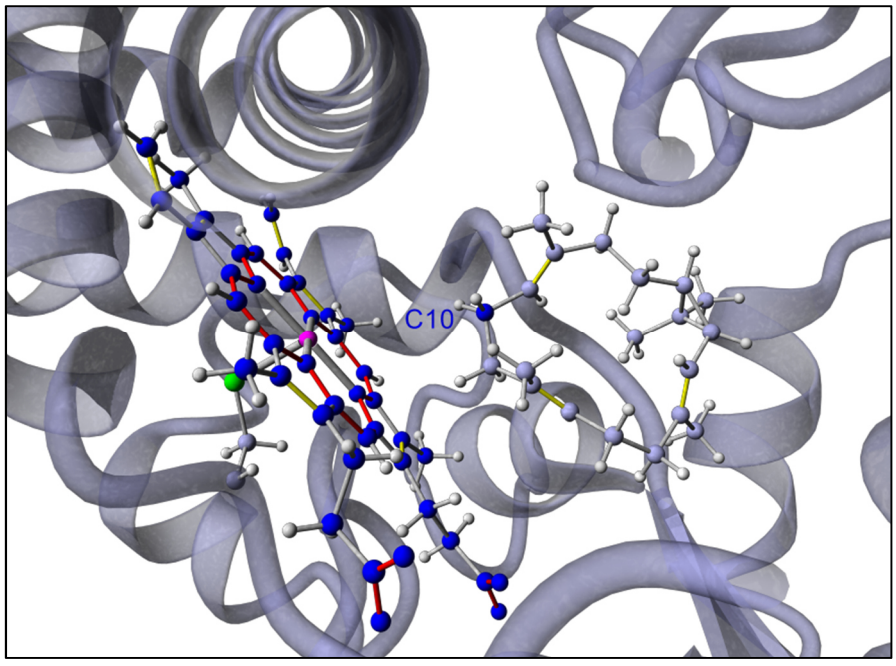


Figure 58. *In silico* modelled CotB3 harboring (-)-casbene. CotB3 (light purple) contains prosthetic heme group (blue) and docked (-)-casbene (gray). The heme group contains the iron cation (magenta). The dative bond to cysteine 408 of the CotB3 is illustrated in green. (-)-casbene is depicted in the conformation comprising the lowest binding energy derived by cluster analysis of AutodockVina.

4.3.7 Characterization and Molecular Adaptation of Class I Redox Systems for the Production of Cyclooctatin

The bioinformatics described in this chapter were done by Patrick Schrepfer and are part of the publication: Görner, C.; Schrepfer, P.; Redai, V.; Wallrapp, F.; Loll, B.; Eisenreich, W.; Haslbeck, M.; Brück, T. *Microb. Cell Fact.* **2016**, *15*, 86.

The redox system AfR/Afx, which allowed successful reconstitution of the cyclooctatin gene cluster was obtained from the genome of *S. afghaniensis*. The bacterial genus *Streptomyces* is known to produce several secondary metabolites including terpenes and polyketides. *S. afghaniensis* has been reported to produce antibiotic sulfur containing polypeptide taitomycin with unknown structure as well as hydroxylated sesquiterpens.[241,242] It can be assumed that AfR and Afx are involved in the hydroxylations of endogenous secondary metabolites. However, the highly efficient functional interaction with CotB3 and CotB4 is surprising, although non-cognate redox partners were shown to activate *Streptomyces*-derived class I P450 monooxygenases.

The molecular interactions involving the electron transfer between ferredoxin and P450 strongly depend on the amino acids decorated at the ferredoxin binding sites. The best characterized molecular recognition and interaction has been examined in the model P450cam•Pdx system.[214] In this model system an induced fit is necessary for the second electron transfer and thus the activity of the P450cam. Consequently, this refined interaction is the reason why P450cam has not been reconstituted by a ferredoxin other than Pdx.[214] Interestingly, this high selectivity is not nessecary for the activation of all class I P450 enzymes.[12] For instance the P450cin, which is involved in the hydroxylation of the monoterpene cineole was demonstrated to have a relaxed redox partner preference.[12] The crystal structure between P450cin and flavoredoxin (Cdx) complex indicated that a structural change is not required.[12]

To elucidate the differences in the redox promiscuity between CotB3, CotB4 and P450cam, the AfR/Afx redox system from *S. afghaniensis* was exchanged with the structurally similar PdR/Pdx system from *P. putida*. To compare different redox system variants with either CotB3 or CotB4, the pACYC-Duet-1 plasmid was changed accordingly. The catalytic efficiency was evaluated using an *in vivo* assay, which was preferred to an *in vitro* assay. In contrast to the *P.putida* derived redox system, the AfR/Afx system displayed very poor protein expression levels thus

challenging a protein purification prior to the application of an *in vitro* assay. To avoid interference with the DXP pathway enzymes, only the pACYC-Duet-1 plasmid was expressed. The metabolic precursors cyclooctat-9-en-7-ol or cyclooctat-9-en-5,7-diol were added to the assay.

For the *in vivo* assay the modified pACYC-Duet-1(AfR/Afx, CotB3), pACYC-Duet-1(AfR/Afx, CotB4), pACYC-Duet-1(PdR/Pdx, CotB3) and pACYC-Duet-1(PdR/Pdx, CotB4) were transformed into *E.coli* Bl21(DE3). As a negative control the empty pACYC-Duet-1 plasmid was used. After a 24 hours protein expression period of the *E.coli* strains harboring the corresponding pACYC-Duet-1 plasmids, the cells' density was normalized followed by the addition of the corresponding metabolic precursor cyclooctat-9-en-7-ol or cyclooctat-9-en-5,7-diol. The *in vivo* assay was performed for 48 hours. The protein expression levels were monitored by taking whole cell SDS samples. At the end of the *in vivo* assay terpenes were extracted, analyzed and the amount of either cyclooctat-9-en-5,7-diol or cyclooctatin was quantified. Interestingly, in an initial experiment it was observed that the selectivity of CotB3 was lost when the metabolic precursor was added from a DMSO stock solution, which led to a final concentration of 1% (*v/v*) DMSO. GC-MS analysis showed the production of a second product with a distinct retention time but comparable MS fragmentation patterns (Figure 59).

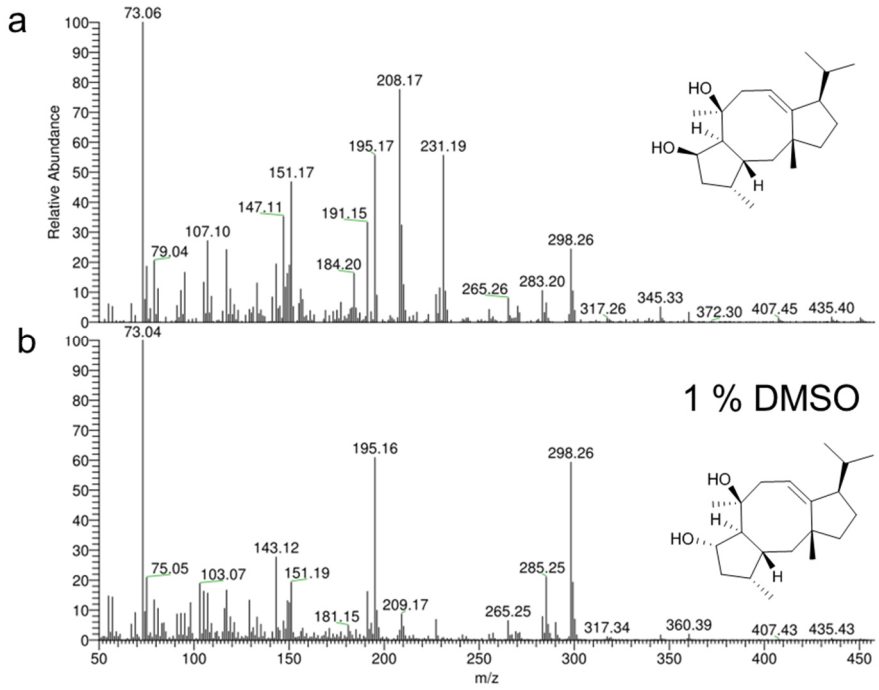


Figure 59. (**a**) MS spectrum of wild type silylated CotB3 product cyclooctat-9-en-5,7-diol. (**b**) MS spectrum of an isomer product of cyclooctat-9-en-5,7-diol isomer, when 1% DMSO (*v/v*) is present.

By contrast, the application of CotB4 under these conditions did not lead to altered product formation. This finding is in agreement with the possible formation of an R and S isomer at C-5 in cyclooctat-9-en-5,7-diol, compared to the formation of a single primary alcohol function at C-18 in cyclooctatin. It was possible to restore the wild type stereoselectivity by applying the metabolic precursors in a acetonitrile stock solution at a final concentration of 0.1% (*v/v*) acetonitrile.

Comparison of the AfR•Afx to the PdR•Pdx redox system showed that in contrast to the AfR•Afx system the PdR•Pdx redox system did not generate detectable amounts of cyclooctat-9-en-5,7-diol with CotB3 (Figure 60a). When the redox systems were evaluated with CotB4, PdR•Pdx gave only minor amounts of cyclooctatin in comparison with AfR•Afx (Figure 60b). SDS-Page analysis of whole cell extracts and subsequent peptide mass fingerprinting was done from cells of the *in vivo* assay with respect to CotB3/4 and the corresponding redox partners (Figure 62/63 below). It was found that in contrast to AfR•Afx the application of the PdR•Pdx system resulted in a significantly increased protein expression of CotB3/4 and Pdx, while the AfR•Afx system showed low expression levels. This finding suggests that the expression of the redox system can significantly influence the expression or stability of the corresponding P450 enzymes.

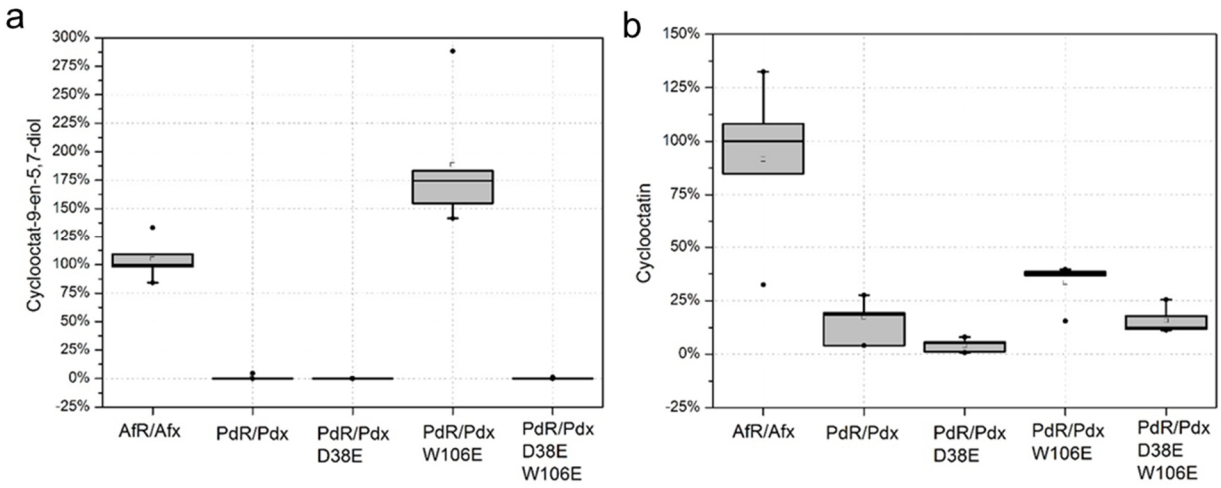


Figure 60. *E.coli* whole cell catalysis experiments. (**a**) *E.coli* cells expressing CotB3 combined with different redox systems and mutants were supplemented with cyclooctat-9-en-7-ol followed by analysis of the cyclooctat-9-en-5,7-diol production yields. (**b**) *E.coli* cells expressing CotB4 combined with different redox systems and mutants were supplemented with cyclooctat-9-en-5,7-diol followed by analysis of the cyclooctatin production yields.

To elucidate the molecular interactions between the AfR•Afx and PdR•Pdx systems with CotB3/B4 an *in silico*-guided mutagenesis approach was applied. In analogy to modelling of CotB3, a chimera model of the crystal structure *S. coelicolor* (PDB 3EL3, 39%, sequence identity) was used as main template and complemented using loop structure refinements from the crystal

structure of human cytochrome P450 (PDB 3LD6, 20% sequence identity). In the next step the models of CotB3 and CotB4 were aligned with the P450cam structure using the reported P450cam•Pdx complex as reference followed by energy minimization and MD simulation. To create a model for the CotB3/4•Afx complexes P450cam and Pdx were replaced in the P450cam•Pdx complex accordingly. The amino acids $D38^{Pdx}$ and $W106^{Pdx}$ were identified to be crucial for the electron transfers in the P450cam•Pdx. Structural superposition of CotB3/4•Afx with P450cam•Pdx showed the structural analogy of $D38^{Pdx}$ with $E39^{Afx}$ as well as $W106^{Pdx}$ with $E107^{Afx}$ (Figure 61).

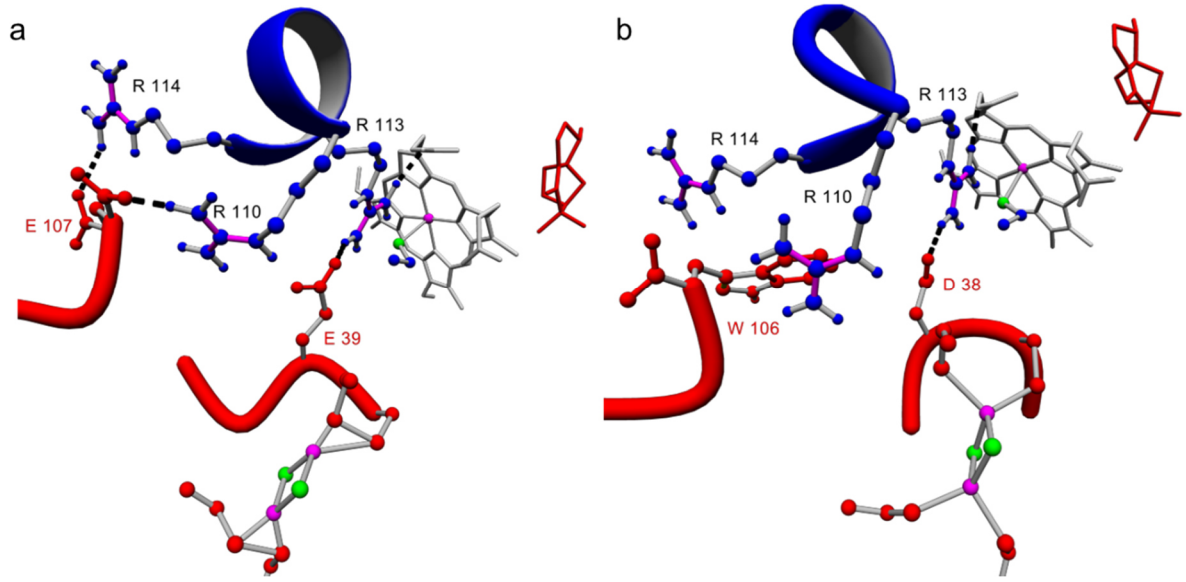


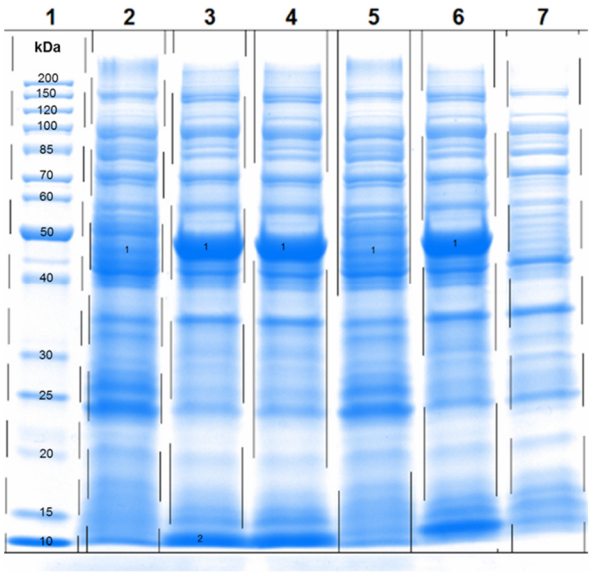
Figure 61. (**a**) Inner view illustrating essential molecular interactions of the *in silico* Cotb3·Afx complex. Afx (red) carries the iron sulfur cluster (green/magenta). An electron is transferred from the iron sulfur cluster to the iron center (magenta) of the heme group (gray) in CotB3 (blue). The substrate cyclooctat-9-en-7-ol (red) located next to the heme group on the opposite side of Afx. Essential hydrogen bonds between Afx and CotB3 are shown in black. (**b**) Inner view illustrating essential molecular interactions of the *in silico* Cotb3·Pdx complex. The graphical presentation is analog to (**a**).

Comparison of the modelled CotB3/4•Afx and CotB3/4•Pdx complexes revealed significant changes in Afx and Pdx interactions (Figure 61). The models suggest the formation of essential hydrogen bonds between $E39^{Afx}$/$R113^{CotB3}$, $E107^{Afx}$/$R110^{CotB3}$, $E107^{Afx}$/$R114^{CotB3}$ and $E39^{Afx}$/$R114^{CotB4}$, $E107^{Afx}$/$R111^{CotB4}$, $E107^{Afx}$/$R115^{CotB4}$. Comparison of these interactions in the CotB3/4•Afx complexes with P450cam•Pdx indicates that the hydrogen bonds between $E39^{Afx}$/$R113^{CotB3}$ and $E39^{Afx}$/$R114^{CotB4}$ are analogous to $E38^{Pdx}$ /$R112^{P450cam}$. Furthermore, the interactions between $E107^{Afx}$/$R110^{CotB3}$ and $E107^{Afx}$/$R111^{CotB4}$ are analogous to $W106^{Pdx}$ /$R109^{P450cam}$. In contrast, the hydrogen bond interactions between $E107^{Afx}$/$R114^{CotB3}$ and $E107^{Afx}$/$R115^{CotB4}$ differ significantly from the hydrophobic interaction between $W106^{Pdx}$/$A113^{P450cam}$. Consequently, in the CotB3/4•Pdx complexes $W106^{Pdx}$ does not display an interaction between $R114^{CotB3}$ and $R115^{CotB}$, respectively. The interaction between $W106^{Pdx}$

and A113^{P450cam} in the P450cam•Pdx complex is essential to initiate a conformational change, which is essential for the second electron transfer.[214] It is likely that the polar interactions between E107^{Afx} and R114^{CotB3}/R115^{CotB4} participate in the binding of the CotB3/4•Afx complex and mediate the second electron transport analogous to the *P. putida* complex. This finding would also explain the significantly increased catalytic efficiency of CotB3/4 when reconstituted with the Afx compared to the Pdx redox system.

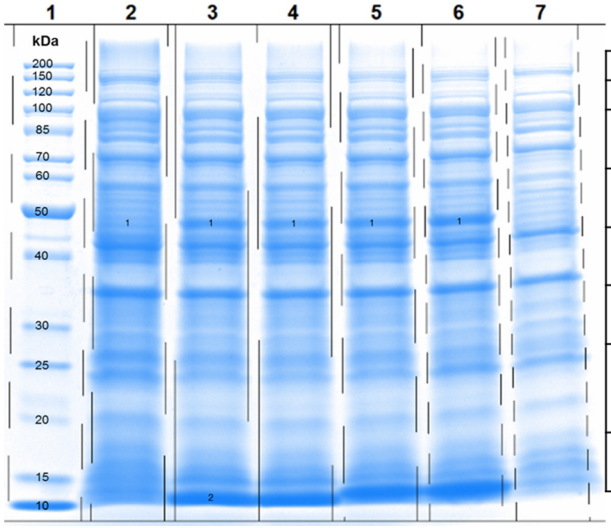
To verify if the contributions of the amino acids E39^{afx} and E107^{Afx} in the Afx ferredoxin are the key to the successful reconstitution of CotB3/B4, the structural analogs D38^{Pdx} and W106^{Pdx} in the Pdx ferredoxin were mutated to the corresponding Afx analogs. Consequently, a series of Pdx mutant variants were created such as Pdx-D38E, Pdx-W106E and Pdx-D38E/W106E. These Pdx mutants were tested for their efficiency to reconstitute CotB3 and CotB4 in the aforementioned *in vivo* assay (Figure 60). Interestingly, the Pdx-W106E mutant showed significantly improved yields of cyclooctat-9-en-5,7-diol compared to the Afx control. In fact the yields were 30% above the Afx reference. Additionally, Pdx-W106E increased the cyclooctatin production but could only reconstitute about 30% of the Afx mediated production yields. In contrast, Pdx-D38E and Pdx-D38E/W106 did not produce any notable amounts of cyclooctat-9-en-5,7-diol and only small quantities of cyclooctatin. While Pdx-D38E decreased the production of cyclooctatin compared to the wild type Pdx, the double mutant Pdx-D38E/W106 produced slightly higher cyclooctatin titers compared to Pdx-D38E.

SDS-PAGE analysis of the whole cell proteome of P450 associated protein complexes derived during the *in vivo* assay (after 24 hours) showed a significant change in the expression levels of the Pdx mutants (Figure 62/63). Interestingly, the expression levels of Pdx-W106E and CotB3 significantly decreased to the level of the CotB3•Afx complex. By contrast, the expression levels of CotB3•Pdx-D38E and CotB3•Pdx-D38E/W106E are very comparable to the wild type CotB3•Pdx complex. The same was true for the CotB4/Pdx-mutants and wild type complexes, which displayed almost identical protein expression levels. Interestingly, when the redox systems AfR/Afx, PdR/Pdx, PdR/Pdx-D38E and PdR/Pdx-W106E were expressed in the absence of CotB3/4, the expression levels of the PdR, Pdx wild type and the Pdx mutant variants were identical. In contrast, the AfR/Afx levels were already significantly decreased and did not distinguish between co-expressions with CotB3/B4.



Lane #	Lane Description
1	Marker
2	pACYC-Duet-1 (afR/afx, cotB3)
3	pACYC-Duet-1 (pdR/pdx, cotB3)
4	pACYC-Duet-1 (pdR/pdx-D38E, cotB3)
5	pACYC-Duet-1 (pdR/pdx-D38E, cotB3)
6	pACYC-Duet-1 (pdR/pdx-D38E/W106E, cotB3)
7	pACYC-Duet-1 (empty) Negative Control

Figure 62. SDS page *in vivo* assay to determine the catalytic efficiency of CotB3 in combination with different redox system variants. Samples were taken after 24 hours and represent the whole cell proteome. Protein band indicated as "1" and "2" was identified by Maldi MS/MS fingerprinting as CotB3 and Pdx, respectively.



Lane #	Lane Description
1	Marker
2	pACYC-Duet-1 (afR/afx, cotB4)
3	pACYC-Duet-1 (pdR/pdx, cotB4)
4	pACYC-Duet-1 (pdR/pdx-D38E, cotB4)
5	pACYC-Duet-1 (pdR/pdx-D38E, cotB4)
6	pACYC-Duet-1 (pdR/pdx-D38E/W106E, cotB4)
7	pACYC-Duet-1 (empty) Negative Control

Figure 63. SDS page *in vivo* assay to determine the catalytic efficiency of CotB4 in combination with different redox system variants. Samples were taken after 24 hours and represent the whole cell proteome. Protein band indicated as "1" and "2" was identified by Maldi MS/MS fingerprinting as CotB4 and Pdx, respectively.

The decreased expression levels of CotB3 in the CotB3•Afx, CotB3•Pdx-W106E complexes as well as CotB4 in the CotB4•Afx complex is remarkable. Since only these variants facilitate strong catalytic activity, a correlation between P450 activity and expression level is very likely. It has been reported that co-expression of P450 and ferredoxin can alter the P450 expression levels.[230] However, in particular the dramatic effect of the single mutation in Pdx-W106E upon the P450 expression level is exceptional. More surprisingly the introduction of a second mutation D38E, restores the P450 and Pdx expression rates to wild type level. These effects

may be explained by oxidative stress caused by activation of CotB3/4, which might ultimately be leading to degradation or reduced expression rates of CotB3/4 and the corresponding ferredoxins. In the case of Afx, the expression rates are very low even in absence of CotB3/4 thus the observed effect is much less pronounced. Contrary to the *in silico* anticipated increase of the catalytic efficiency, Pdx-D38E did not increase in efficiency for CotB3 or CotB4. $D38^{Pdx}$ has been reported to play an important part in the electron transport between PdR and Pdx as well as between Pdx and P450cam.[12,214] Even though the exchange from Asp to Glu only extends the side chain by a CH_2 group, this subtle change may significantly alter the efficiency of the electron transfer. In fact, Pdx-D38E may lead to a more effective binding to CotB3/B4 as predicted by the *in silico* modelling. However, at the same time the electron transfer between PdR and Pdx-D38E is disturbed and higher catalytic activity of the P450s are not observed.

The moderate production improvements observed in the CotB4•Pdx-W106E complex compared to CotB4•Afx, indicate a less efficient activation of CotB4 by Pdx-W106E compared to CotB3. Consequently, it can be assumed that the binding modes of CotB3 and CotB4 with Afx are more complex and cannot be reduced to interaction of $E107^{Afx}$. For further optimizations other residues need to be taken into account. Nevertheless, this research has shown that $E107^{Afx}$ is a top candidate for further optimization. Other optimizations should focus on increasing the binding affinity between P450 and ferredoxin by hydrogen bonds and van der Waals' interactions. To identify further amino acid interaction crucial for the electron transfer, a high throughput mutagenesis and screening approach is required to enable early identification of negative protein alternations.

4.4 Conclusion and Outlook

The identification of the AfR•Afx redox system was essential to reconstitute the CotB3/4 hydroxylase activity, which ultimately allowed biotechnological production of cyclooctatin in an engineered *E.coli* strain. Since the application of non-native redox system often results in low yields, the newly identified AfR•Afx redox system may be used for the pathway reconstitution of other *Streptomyces* derived from secondary product metabolism in *E.coli*. Furthermore, it has been shown that *in silico* identification key amino acid such as the W106 in the putida ferredoxin can be used to find suitable variants that participate in the P450 ferredoxin interaction. Consequently, the mutagenesis experiments used in this survey clearly demonstrated that by a single amino acid exchange a barley active non-native redox system such as the PdR•Pdx system from *P. putida* can be engineered to efficiently reconstitute class I P450 monooxygenases. This finding may be particularly useful to reconstitute P450s with redox systems from phylogenetically different strains. The successful reconstitution of CotB3 led to the discovery of the first biotechnological production of the hydroxylated casbene diterpene sinularcasbane D. Due to the substrate promiscuity, CotB3 may be a valuable engineering target for the production of other native and non-native functionalized diterpenes. Since only very few diterpene hydroxylases have been reconstituted, CotB3 as a high active P450 monooxygenase is of particular interest.

5. Part II & III: Materials and Methods

General Experimental Procedures

Chemicals used in this study were obtained from Sigma Aldrich at the highest purity grade available. NMR spectra were recorded in $CDCl_3$ or CD_3OD with an Avance-III 500 MHz (Bruker) at 300 K. ^1H NMR chemical shifts are given in ppm relative to $CHCl_3$ (δ=7.26 ppm) or CD_2HOD (δ=3.31 ppm) and CD_3OH (δ=4.87) (^1H NMR). ^{13}C NMR chemical shifts are given in ppm relative to $CDCl_3$ at δ=77.16 ppm or CD_3OH at δ=49.00 ppm. The 2D experiments (HSQC, COSY and NOESY) were performed using standard Bruker pulse sequences and parameters.

GC-MS/FID Sample Preparation

GC-MS analysis of diterpene products from ethyl acetate extractions was done by a Trace GC Ultra with DSQII (Thermo Scientific). CotB2 mutant diterpene products were analyzed without derivatization. For the analysis of CotB3 and CotB4 metabolic product probes were silylated prior to measurement. Therefore, solvents of diterpene solutions were removed to dryness under nitrogen flow. The residue was solved in a mixture of 37.5% (*v/v*) N,O-bis(trimethylsilyl)trifluoroacetamide (BSTFA) and 25% (*v/v*) chlorotrimethylsilan (TMS-Cl) in pyridine. The derivatization reaction was done at 70°C for 1 hour.

Derivatization of terpenes solved in dodecane was done without evaporating the solvent. Therefore, 50 µL of the dodecane probes were mixed with 45 µL pyridine, 45 µL N,O-bis(trimethylsilyl)trifluoroacetamide (BSTFA) and 10 µL (*v/v*) chlorotrimethylsilan (TMS-Cl). Silylation was carried out at 70 °C for 4 hours.

GC-MS/FID Analysis

One µl of the sample was applied by a TriPlus Autosampler onto a SGE BPX5 column (30 m, I.D 0.25 mm, Film 0.25 µm). The initial column temperature started at 50°C and was maintained for 2.5 min. A temperature gradient was applied from 50°C – 320°C (10°C/min), followed by 3 min maintenance at 320°C. MS data was recorded at 70 eV (EI), *m/z* (rel. intensity in %) as TIC, total ion current. The recorded *m/z* range covered 50 – 650. Mass-spectrometric fragmentation patterns were analyzed with the NIST NIST08MS spectral database library, which was extended by a special natural product database.[182]

Quantification was done by using flame ionization detector (FID) using 1 mg/mL α-humulene (Sigma-Aldrich, Germany) as internal standard. To apply the FID signal of α-humulene to the silylated diterpene products, correlation factors were calculated as previously described.[243] After the concentration of the silylated terpene species were determined, the concentration of the non-derivate terpenes was calculated.

High Resolution MS

High-resolution mass spectra of diterpenes was determined with a Thermo Fisher Acela HPLC system linked to a Thermo Fisher Scientific LTQ Orbitrap XL mass. Electrospray ionization was done in positive ion mode.

Proteomics

12% SDS-Page gels according to Lämmili were used for SDS-PAGE analysis. To identify recombinantly expressed proteins as protein bands on SDS-PAGE, the bands were excised and digested with trypsin as described previously.[244] The peptides were concentrated and purified using C18 ZipTips (Merck Millipore) according to the manufacturer's protocol. The peptides were analyzed by MALDI-MS using an Ultraflex I (Bruker Daltronics). Data analysis was performed using the MASCOT software program (Matrix Science, London, UK) along with the National Center for Biotechnology Information database.

Circular Dichroism (CD) Spectroscopy and Optical Rotation

Circular dichroism (CD) spectroscopy was performed using a Chirascan plus spectropolarimeter (Applied Photophysics). Samples were dissolved in acetonitrile and spectra were recorded in quartz cuvettes with 0.1 cm path length at 20°C. Optical rotations were measured using a PerkinElmer 241MC polarimeter

Bacterial Strains, Genes and Vectors

The *E.coli* strains XL1-Blue ((recA1, endA1, gyrA96, thi-1, hsdR17, supE44, relA1, lac[F′proAB $lacI^qZ\Delta M15Tn10(Tet^r)$]) was used for cloning, HMS174(DE3) (F^- *recA1 hsdR*(r_{K12}^- m_{K12}^+) (DE3) (Rif ^{R})) was used for the *in vitro* screening of CotB2 mutants and Bl21 (DE3) (F– *omp*T *hsd*SB(rB–, mB–) *gal dcm* (DE3)) was used for *in vivo* production of the diterpenes. *E.coli* strains and the plasmids pET24a(+), pET-Duet-1, pACYCDuet-1, pCDFDuet-1 and pCOLADuet-1 were obtained from Novagen/ Merk Millipore. PCR primers were synthesized by Eurofins MWG Operon. All genes used for cloning were synthesized by Life technologies GmbH containing the appropriate restriction sites and adjusting codon usage for *E.coli*. and are shown in the Appendix (Figure 81-90). One exception was the geranylgeranyl diphosphate synthase CrtE

(GenBank: M90698.1), which was amplified from the bacteria *Pantoea agglomerans* (ATCC 27155) using standard protocols. Primers used were: 5'-AAA CCA TGG CAA TGG CAA CGG TCT GCG CA-3' and 5'-AAA GAA TTC TTA ACT GAC GGC AGC GAG TTT-3'.

Plasmids for Targeted Engineering of Cyclooctat-9-en-7-ol Synthase CotB2

A *E.coli* optimized gene of cotB2 (GenBank: BAI44338.1) was cloned into the plasmid pET24(+) using the NdeI and XhoI restriction site by standard cloning techniques. Mutagenesis was done according to the QuickChange Site-directed mutagenesis protocol. Primers used for mutagenesis are shown in Table 7.

Table 7. Primers used for the mutagenesis of cotb2.

Primer	Sequence (5' -> 3')	Annealing temperature
F107X_F	GTG TGC GTT CTG AAT TGT GTT ACC NNN GTT TGG GAT GAT ATG GAC CCT GC	70°C
F107X_R	GCA GGG TCC ATA TCA TCC CAA CAN NNG GTA ACA CAA TTC AGA ACG CAC AC	70°C
F149X_F	GTT GCC TAT GAA GCA GCA CGT GCA NNN GTT ACC AGC GAT CAC ATG TTT CG	70°C
F149X_R	CGA AAC ATA GTG ATC GCT GGT AAC NNN TGC ACG TGC TGC TTC ATA GGC AAC	70°C
W186X_F	GTT ACC GAT ATT GGC GTG GAT TTT NNN ATG AAA ATG AGC TAT CCG ATT	55°C
W186X_R	AAT CGG ATA GCT CAT TTT CAT NNN AAA ATC CAC GCC AAT ATC GGT AAC	55°C
N285X_F	GTT TTT CTG GAT CTG ATT TAT GGC NNN TTT GTG TGG ACC ACC TCC AAC AA	55°C
N285X_R	TTG TTG GAG GTG GTC CAC ACA AAN NNG CCA TAA ATC AGA TCC AGA AAA AC	55°C
W288X_F	CTG ATT TAT GGC AAT TTT GTG NNN ACC ACC TCC AAC AAA CGT TAT AAA AC	55°C
W288X_R	GTT TTA TAA CGT TTG TTG GAG GTG GTG CCC ACA AAA TTG CCA TAA ATC AG	55°C
Y77X_F	TTA GTG ATG AAC GTT GGA TTA GCN NNG TTG GTG TTG TTC TGT GGT CA	55°C
Y77X_R	TGA CCA CAG AAC AAC ACC AAC NNN GCT AAT CCA ACG TTC ATC ACT AA	55°C
Y295X_F	GAC CAC CTC CAA CAA ACG TNN NAA AAC CGC AGT GAA TGA TG	55°C
Y295X_R	CAT CAT TCA CTG CGG TTT TNN NAC GTT TGT TGG AGG TGG TC	55°C
N110E_F	GTT ACC TTT GTT TGG GAA GAT ATG GAC CCT GCA C	55°C
N110E_R	GTG CAG GGT CCA TAT CTT CCC AAA CAA AGG TAA C	55°C
N111E_F	CTT TGT TTG GGA TGA AAT GGA CCC TGC ACT G	55°C
N111E_R	CAG TGC AGG GTC CAT TTC ATC CCA AAC AAA G	55°C
N113E_F	CTT TGT TTG GGA TGA TAT GGA ACC TGC ACT GCA TGA TTT TGG	55°C
N113E_R	CCA AAA TCA TGC AGT GCA GGT TCC ATA TCA TCC CAA ACA AAG	55°C

W288G_F	CTG ATT TAT GGC AAT TTT GTG GGC ACC ACC TCC AAC AAA CGT TAT AAA AC	55°C
W288G_R	GTT TTA TAA CGT TTG TTG GAG GTG GTG CCC ACA AAA TTG CCA TAA ATC AG	55°C
Y107A_F	GTG TGC GTT CTG AAT TGT GTT ACC GCG GTT TGG GAT GAT ATG GAC CCT GC	55°C
Y107A_R	GCA GGG TCC ATA TCA TCC CAA ACC GCG GTA ACA CAA TTC AGA ACG CAC AC	55°C

Saturation mutagenesis of F107, F149, W186, N285 was done using the PfuTurbo Cx Hotstart DNA Polymerase from Agilent Technologies (Santa Claram, CA, USA). All other mutagenesis PCRs were executed with the Phusion High Fidelity DNA Polymerase (Thermo Scientific, Rockford, Illinois, USA). PCR reactions were done in 50 µL volumes supplemented with 1- 10 ng template plasmid, 10 pmol phosphorylated primers, 0.2 mM of each deoxynucleotide triphosphate (dNTPs) and 1.25 U PfuTurbo Cx Hotstart DNA Polymerase or 2.5 U Phusion High Fidelity DNA Polymerase. For the PfuTurbo Cx Hotstart DNA Polymerase thermal cycling was set up in the following way:

A 3 minutes heating step at 95°C (denaturation), followed by 30 cycles of 45 sec at 95°C (denaturation), 30 sec at 70°C (annealing) and 10 minutes 72°C (elongation). After the cycling ended a 15 minutes 72°C elongation step was attached. After PCR reaction was completed the reaction mixture was stored at 16°C. The program for the Phusion High Fidelity DNA Polymerase was done as follows. A 2 minutes 98°C denaturation step, followed by 30 cycles of 20 seconds 98°C (denaturation), 45 sec 55°C (annealing) and 30 seconds/kb (elongation). After that a 10 minutes 68°C step was added. After completion the PCR reaction mixture was stored at 16°C.

After the PCR reaction, the methylated DNA was digested by the addition of 30 U DPN1 (Thermo Scientific, Rockford, Illinois, USA). The reaction mixture was incubated at 37°C for 3 hours. Purification was done using the GeneJET Plasmid Miniprep Kit (Thermo Scientific, Braunschweig), according to manufacturer's protocol and transformed into chemical competent *E.coli* XL1-Blue cells as described elsewhere.[125] Plasmids from *E.coli* XL1-Blue were extracted and purified using the GeneJET Plasmid Miniprep Kit (Thermo Scientific, Braunschweig), according to manufacturer's protocol. All clones were validated by DNA sequencing.

In Vitro Screening of CotB2 Mutants

The gene libraries were transformed into chemically competent *E.coli* HMS174(DE3) cells using standard transformation protocols. Cultivation was done in 50 mL LB-Medium supplemented with 50 µg mL^{-1} kanamycin. The cells were grown to an OD_{600} 0.6. A volume of 1mL was taken from every culture and was used for plasmid extraction (GeneJET Plasmid Miniprep Kit (Thermo Scientific, Braunschweig)) following manufacturer's protocol. Protein expression was initiated by the addition of 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG) in the cultures and was carried out over night at 25°C. The cells were harvested by pelleting (10 min, 4500 rpm, 4°C) followed by a washing step with Tris-buffer (50 mM, pH = 7.5). The cells were pelleted again (10 min, 4500 rpm, 4°C) and resuspended in the assay buffer (50 mM Tris-HCL, 1 mM $MgCl_2$, pH = 7.5). In the next step, the cells were lysed by sonification using a Sonoplus HD2070 (Bandelin Electronic) performing 3 repeats on ice (45 sec on and 3 min off at 60% power). The cell lysate was centrifuged (20 min, 15000 rpm, 4°C) and 125 µL of the supernatant was diluted with 125 µL assay buffer (50 mM Tris-HCL, 1 mM $MgCl_2$, pH = 7.5). The reaction was started by the addition of GGDP (Sigma-Aldrich) to a final concentration of 45 µM. The assay was incubated at 29°C shaking at 350 rpm overnight. Diterpene products were extracted with 1 mL ethyl acetate followed by organic phase separation and drying by $MgSO_4$. The organic solvent was reduced to 50 µL under nitrogen flow and analyzed by GC-MS.

E.coli Based Production of Cyclooctat-7-en-9-ol and Cyclooctat-1,7-diene

For the production of the CotB2 mutants F107A and F149L, an *E.coli* diterpene production system was created with minor modifications as previously described. Therefore, the plasmid pCDFDuet-1 harboring a codon optimized 1-deoxyxylulose-5-phosphate synthase (dxs) (GenBank: YP_001461602.1) in the first multiple cloning side (NdeI, XhoI) and a codon optimized bi-cistronic operon of the isopentenyl diphosphate isomerase (idi) (GenBank: NP_417365.1) and 1-deoxy-D-xylulose-5-phosphate reductoisomerase (dxr) (GenBank: NP_414715.1) in the second multiple cloning side (NdeI and XhoI) was created. Furthermore the geranylgeranyl diphosphate synthase (crtE) (GenBank: M90698.1) was introduced into the second multiple cloning side (NdeI and XhoI) of the pACYCDuet-1 vector. Both plasmids and the plasmid pET24a(+) harboring the mutant CotB2 variants were introduced into *E.Coli* Bl21(DE3) by standard transformation methods.

The cultivation conditions were based on Morrone *et al.* 2010. Cultures were grown in three times 2 L shake flakes containing FL-medium (10 g/L yeast extract, 20 g/L casein hydrolysate, 5 g/L NaCl, 1 g $MgSO_4$) supplemented with 50 µg mL^{-1} kanamycin, 34 µg mL^{-1} chloramphenicol and 50 µg mL^{-1} streptomycin at 37°C. The cultures were shaken at 200 rpm and grown to an OD_{600} = 0.6. Next, the cultures were cooled down to 16°C for 1 hour followed by the addition of 5 mL 1M phosphate buffer (pH 7) and glucose to a concentration of 2% (*w/v*). The cultures were incubated for another hour at 16 °C. Protein expression was initiated by the addition of IPTG (1 mM) and the cells were incubated for 60 hours at 16 °C. After 36 hours 2 mM pyruvate was added in 12 hour intervals.

To isolate terpene products, the cells were pelleted (15 min, 10000 rpm, 4°C) and the supernatant was extracted by ethyl acetate (200 mL/L fermentation medium). The organic phases were combined and dried by $MgSO_4$ and concentrated. Purification was carried out in two steps. The first step comprised a reverse phase chromatographic step with a water acetonitrile gradient on a (Polygoprep 60-50, C_{18}, Macherey-Nagel) chromatography flash column. The second step was a normal phase chromatographic step by a pentane/ethyl acetate gradient on a silica phase column (Silica gel 40, Fluka Analytical).

In vivo

Screening of CotB2 Mutants

The plasmid based DXP pathway used for the *in vivo* system was based on the plasmids pCola-Duet-1, pCDFDuet-1 and pET-Duet-1 harboring the DXP pathway genes and the diterpene synthase CotB2. Codon optimized genes of the 1-deoxy-D-xylulose 5-phosphate synthase (*dxs*) (GenBank: YP_001461602.1), 1-deoxy-D-xylulose 5-phosphate reductoisomerase (*dxr*) (GenBank: NP_414715.1), 2-C-methyl-D-erythriol 4-phosphate cytidyltransferase synthase (*ispD*) (GenBank: NP_417227.1), 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (*ispF*) (GenBank: NP_289295.1) and isopentenyl-diphosphate delta isomerase (*idi*) (GenBank: NP_417365.1) were synthesized. *IspD/ispF* was synthesized as a bi-cistronic operon. The gene for the geranylgeranyl diphosphate synthase (*crtE*) (GenBank: M90698.1) was isolated as aforementioned. The plasmid assembly is shown in table 8 and was done using standard cloning techniques.

Table 8. Plasmids used to construct the overexpressed DXP pathway in *E.coli* Bl21 (DE3).

Gene(s)	Vector	Multiple Cloning Site	Restriction Sites
dxr	pCola-Duet-1	I	NcoI, EcoRI
dxs	pCola-Duet-1	II	NdeI, XhoI
ispD/ispF operon	pCDF-Duet-1	I	NcoI, EcoRI
idi	pCDF-Duet-1	II	NdeI, XhoI
crtE	pET-Duet-1	I	NcoI, EcoRI
cotB2	pET-Duet-1	II	NdeI, XhoI

The *in vivo* screening of diterpene mutants was done in shake flask containing 10 mL M9 minimal medium supplemented with 30 µg mL^{-1} kanamycin, 50 µg mL^{-1} ampicillin and 50 µg mL^{-1} streptomycin. Therefore, 1 mL of an overnight culture in LB medium (30 µg mL^{-1} of kanamycin, 50 µg mL^{-1} ampicillin and 50 µg mL^{-1} streptomycin) was washed in M9 minimal medium and used for inoculum. The cells were cultured at 37°C to an OD_{600} of 0.6 followed by supplementation of 1% (*w*/*v*) POLYGROPREP® 60-50 C_{18} Beads, 2% glycerol (*v*/*v*), 1 g/L casamino acids and 1 mM IPTG. Cultivation was continued at 25°C for 20 hours.

After the cultivation the C_{18} beads were filtered and washed with ddH_2O. Terpenes were extracted from the hydrophobic beads by using ethyl acetate. Therefore the beads were suspended in ethyl acetate followed by vigorous mixing. The organic phase was separated by centrifugation. Ethyl acetate extraction of the C_{18} beads was repeated 3-4 times. The organic phases were combined, dried by $MgSO_4$ and the solvent was removed under vacuum. The crude extract was solved in 1 mL ethyl acetate and analyzed by GC-MS.

Production of (1R,3E,7E,11S,12S)-3,7,18-dolabellatriene and R-cembrene A

The diterpene products (1R,3E,7E,11S,12S)-3,7,18-dolabellatriene and R-cembrene A were produced by the mutants CotB2 W288G and CotB2 F107A, respectively. For production of diterpenes in the mg scale, the *in vivo assay* used to screen CotB2 mutants was upscaled. Therefore cultivation was performed in a 1 L M9 medium supplemented with 1 g/L casamino acids, 30 µg/mL kanamycin, 50 µg/mL streptomycin, 50 µg/mL and carbenicillin in 5 L baffled glass flasks. Cultivation was inoculated at OD_{600} 0.1 (overnight culture: 8 h cultivation at 37 °C, LB medium supplemented with 30 µg/mL kanamycin, 50 µg/mL, streptomycin and 50 µg/mL carbenicillin) and grown to an OD_{600} 0.8 at 37 °C. Subsequently, 0.5 mM isopropyl β-D-1-

thiogalactopyranoside (IPTG), 2 (*w/w*) glycerol and 2.5 g/L hydrophobic beads (Polygoprep 60-50 C18, Macherey-Nagel) were added to the cultivation medium. Cultivation was performed at 28 °C for 5 days.

After 3 days an additional 2% (*w/w*) of glycerol was added. To isolate diterpene products from the hydrophobic beads, the hydrophobic beads were recovered by filtration. The beads were washed with ddH_20 and extracted 3 times by 200 mL ethyl acetate. Organic phases were combined and dried by $MgSO_4$. The organic solvent was removed under vacuum. For purification, the crude extract was solved in 1 mL ethyl acetate and applied to a normal phase flash chromatography purification step using hexane on silica (Silica gel 40, Sigma-Aldrich) as solid phase.

Production of Cyclooctatin

For the production of cyclooctatin the *in vivo* screening system as described in "*In vivo* screening of CotB2 mutants" was extended by the plasmid pACYC-Duet-1. Bicycstronic operons of the P450 monooxygenases *cotB3* (GenBank: BAI44339.1) and *cotB4* (GenBank: BAI44340.1) were created. Furthermore operons of the reductase (*afR*) (GenBank: WP_020277402) and ferrodoxin (*adx*) (GenBank: WP_020276845) from *Streptomyces afghaniensis* were synthesized. The bi-cystronic operons were introduced into the plasmid pACYC-Duet-1 as described on table 9 by standard cloning techniques.

An *E.coli* Bl21 (DE3) strain harboring the plasmids pCola-Duet-1 (*dxp*, *dxs*), pCDF-Duet-1 (*ispD*/*ispF*, *idi*), pET-Duet-1 (*crte*, *cotB2*) and the vector pACYC-Duet-1 (*afR*/*afx*, cotB3/cotB4) was created by standard transformation techniques. Cultivation conditions were based on Boghigian *et al.*[237]

Table 9. Plasmids used to construct the cyclooctatin biosynthesis in *E.coli* Bl21 (DE3).

Gene(s)	Vector	Multiple Cloning Site	Restriction Sites
afR/afx operon	pACYC-Duet-1	I	NcoI, NotI
cotB3/cotB4 operon	pACYC-Duet-1	II	NdeI, XhoI

Shake-flask Production Culture of Cyclooctatin

Cyclooctatin production in shake flask cultures was done by cultivation of 400 mL LB medium (10 g/L tryptone, 5 g/L yeast extract, and 10 g/L NaCl) in 5 L baffled glass flasks supplemented with 100 mM HEPES (pH 7.6), 40 g/L glycerol, 30 µg/mL kanamycin, 50 µg/mL streptomycin, 50 µg/mL carbenicillin, 34 µg/mL chloramphenicol, 1 mM δ-aminolevulinic acid, 1 mM $FeSO_4$ x

$7H_2O$, 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and 40 µL 30% Antifoam A (Sigma Aldrich). The culture was inoculated at OD_{600} 0.1 from an overnight culture (8 h cultivation at 37°C supplemented with 30 µg/mL kanamycin 50 µg/mL streptomycin, 50 µg/mL carbenicillin, 34 µg/mL chloramphenicol) and cultivated at 25°C for 4 days.

Batch Bioprocess of Cyclooctatin

A 5L fermentation was performed in a 10 L bioreactor (Biostat C, Braun Melsungen, Germany) using LB-medium (10 g/L tryptone, 5 g/L yeast extract, and 10 g/L NaCl) supplemented with 40 g/L glycerol, 30 µg/mL kanamycin, 50 µg/mL streptomycin, 50 µg/mL carbenicillin, 34 µg/mL chloramphenicol, 1 mM δ-aminolevulinic acid, 1 mM $FeSO_4$ x $7H_2O$ and 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). The pH was controlled at 7.6 with 4M NH_4OH and 5 M H_3PO_4. For the inoculation of the fermenter, biomass of an overnight culture (8 h cultivation 37°C supplemented with 50 µg/mL kanamycin, 50 µg/mL streptomycin, 50 µg/mL carbenicillin, 34 µg/mL chloramphenicol) was added to yield an OD_{600} 0.1. Fermentation was run for 4 days at 25°C maintaining an oxygen at 80%. To monitor the diterpene production, 200 mL aliquots were taken every 24 hours in triplicates. To determine the dry cell mass, 10 mL aliquots were taken, centrifuged and dried. The biomass was determined gravimetrically.

Batch Bioprocess of Cyclooctatin with Organic Overlay

A 1L fermentation was performed in a 3 parallel bioreactors (Labfors 5 Lux Stirred Tank, Infors HT, Switzerland). The fermentation medium was LB-medium (10 g/L tryptone, 5 g/L yeast extract, and 10 g/L NaCl) supplemented with 40 g/L glycerol, 30 µg/mL kanamycin, 50 µg/mL streptomycin, 50 µg/mL carbenicillin, 34 µg/mL chloramphenicol, 1 mM δ-aminolevulinic acid , 1 mM $FeSO_4$ x $7H_2O$ and 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). 200 mL sterile dodecane was added to the fermentation medium. The pH was controlled at 7.6 with 4M NH_4OH and 5 M H_3PO4. Inoculation of the bioprocess was done from an overnight culture (8 h cultivation 37°C supplemented with 50 µg/mL kanamycin, 50 µg/mL streptomycin, 50 µg/mL carbenicillin, 34 µg/mL chloramphenicol) at OD600 0.1. The fermentation was executed for 4 days at 25°C. Oxygen was supplied at 0.9 (vvm) and oxygen saturation was constantly adjusted to 80%. The diterpene production was monitored by taking 1 ml aliquots of the dodecane layer every 24 hours. In order to monitor the dry cell mass, 2 mL aliquots were taken, centrifuged and dried. The biomass was determined gravimetrically.

Extraction and Isolation of Cyclooctatin and its Biosynthetic Precursors

Cyclooctatin and its biosynthetic precursors were isolated from shake flask cultures or bioprocess cultures without a dodecane overlay. Product isolation was done from the *E.coli* cell pellet (centrifugation: 15 min, 17,500 g, 4°C) and the growth medium supernatant in separate approaches. The cell pellet was washed with water, resuspended in 5 mL water and sonificated using a Sonoplus HD2070 (Bandelin Electronic) performing 5 repeats on ice (5 min on and 3 min off at 80% power). The lysate was extracted 3 times by using 25 mL ethyl acetate. Furthermore, the supernatant was extracted 3 times by using 200 mL ethyl acetate. All organic phases were combined and dried by $MgSO_4$. The organic solvent was removed under vacuum. The crude extract was solved in 1 mL ethyl acetate and analyzed by GC-MS and -FID. Purification of cyclooctat-9-en-7-ol was carried out by flash chromatography. Therefore an isocratic 70/30 hexane/ethyl acetate silica step (Silica gel 40, Sigma-Aldrich,) was done followed by an isocratic 30/70 water/acetonitrile reversed phase chromatography step on a Polygoprep 60-50, C_{18} column (Macherey-Nagel). For purification of cyclooctat-9-en-5,7-diol by flash chromatography, the hexane/ethyl acetate solvent was changed to 50/50 and the water/acetonitrile solvent to 10/90. Cyclooctatin was purified by flash chromatography using hexane/ethyl acetate 30/70 as solvent followed by 10/90 water acetonitrile solvent in the second purification step.

Plasmids Used to Evaluate Hydroxylation of Diterpenes by the P450 Monooxygenases CotB3 and CotB4

To evaluate the ability of the P450 hydroxylases CotB3/4 to hydroxylate non-native diterpene macrocycles, the diterpene synthase gene *cotb2* in the pET-Duet-1 plasmid was exchanged with different diterpene syntases or mutant variants of CotB2 according to table 10. Cloning was analog as described in the case of CotB2.

Table 10. Plasmids used to evaluate hydroxylation of diterpenes by CotB3/4

Gene(s)	Description	Diterpene produced
cs	Casbene synthase	(-)- Casbene
txs	Taxadiene synthase	Taxa-4,11-diene
$cotb2_{W288G}$	Cyclooctatin synthase	(1R,3E,7E,11S,12S)-3,7,18-Dolabellatriene
$cotb2_{F107A}$	Cyclooctatin synthase	(R)- Cembrene A

Evaluating CotB3/4 for Activity on Different Diterpene Skeletons

The vectors pCola-Duet-1 (*dxp*, *dxs*), pCDF-Duet-1 (*ispD*/*ispF*, *idi*), pET-Duet-1 (*crte*, *cs*)/ pET-Duet-1 (*crte*, *txs*)/ pET-Duet-1 (*crte*, $cotb2_{W288G}$) / pET-Duet-1 (crte, $cotb2_{F107A}$) and the vector pACYC-Duet-1 (afR/afx, cotB3)/ pACYC-Duet-1 (afR/afx, cotB4) were introduced into *E.coli* Bl21 (DE3) by standard transformation procedures. Cultivation was done in shake flaks analog as described in the batch bioprocess of cyclooctatin.

Batch Bioprocess of Sinularcasbane D

An *E.coli* Bl21 (DE3) strain harboring the plasmids pCola-Duet-1 (*dxp*, *dxs*), pCDF-Duet-1 (*ispD*/*ispF*, *idi*), pET-Duet-1 (*crte*, *cs*) and the vector pACYC-Duet-1 (*afR*/*afx*, *cotB3*) was created by standard transformation procedures. Cultivation conditions were based on Boghigian *et al.*[237] The cells were grown in 5 L baffled shake flasks containing 400 mL LB-medium (10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl), 40 g/L glycerol, 30 µg/mL kanamycin, 50 µg/mL streptomycin, 50 µg/mL carbenicillin, 34 µg/mL chloramphenicol, 1 mM δ-aminolevulinic acid , 1 mM $FeSO_4$ x $7H_2O$, 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and 40 µL 30% Antifoam A. The culture was inoculated at OD_{600} 0.1 from an overnight culture (8 h cultivation 37°C supplemented with 30 µg/mL kanamycin 50 µg/mL streptomycin, 50 µg/mL carbenicillin, 34 µg/mL chloramphenicol). Cultivation was done for 3 days at 25°C.

Extraction and Isolation of (-)-Casbene and Sinularcasbane D

Since only 20% of (-)-casbene were converted by the monooxygenase CotB3 to sinularcasbane D, (-)-casbene and sinularcasbane D were purified from the same culture. Harvesting and terpene extraction was done as described for cyclooctatin and its biosynthetic precursors. After the removal of the organic solvent, the crude extract was solved in 1 mL ethyl acetate and analyzed by GC-MS and -FID. Purification of (-)-casbene and sinularcasbane D was carried out by flash chromatography. Therefore an isocratic 90/10 hexane/ethyl acetate chromatography on silica (Silica gel 40, Sigma-Aldrich) was sufficient.

Preparation and Purification of MTPA Derivatives

5 mg purified cyclooctat-9-en-5,7-diol was esterified with the R- and S- Mosher acid according to protocol.[236] The reaction mixtures were purified by flash chromatography using an isocratic 50/50 hexane/ethyl acetate silica phase column (Silica gel 40, Fluka Analytical).

Plasmids to Evaluate Different Redox-System Variants

The catalytic activity of the individual P450 monoxygenases CotB3 and CotB4 proteins was evaluated using different redox-system variants. Tested variants comprised the redox system

of *S. afghaniensis* (afR/afx), *P. putida* (pdR/pdx) and *P. putida* (pdR/pdx) with mutations in the ferredoxin Pdx. Therefore, multiple variants of the pACYC-Duet-1 plasmid were created carrying either CotB3 or CotB4 in multiple cloning site II as well as the bi-cistronic operon of reductase/ ferredoxin in the multiple cloning site I. Different pACYC-Duet-1 plasmid created in this study are shown in table 11.

Table 11. Plasmids used to compare the activity of CotB3 and CotB4 hydroxylases with different redox system variants.

P450 Proteins	Redox system	Mutations
CotB3	AfR/Afx	-
CotB3	PdR/Pdx	-
CotB3	PdR/Pdx	D38E
CotB3	PdR/Pdx	W106E
CotB3	PdR/Pdx	D38E, W106E
CotB4	AfR/Afx	-
CotB4	PdR/Pdx	-
CotB4	PdR/Pdx	D38E
CotB4	PdR/Pdx	W106E
CotB4	PdR/Pdx	D38E, W106E

The *cotb3* gene was copied from the bi-cistronic operon (*cotB3*/*cotB4*) by PCR using the primers 5′-ATT ACA TAT GCG TGA ACG TGG T-3′ and 5′-ATA TCT CGA GTT AAC GCG GTT CAC AAA CCA. The annealing temperature was 55°C and the PCR was done using standard conditions. *CotB4* was amplified from the bi-cistronic operon (cotB3/cotB4) using the primers: 5′-ATT ACA TAT GAA AGA TTT TTT TCG TAT GCG CAC-3′ and 5′-ATA TCT CGA GTT AAC GAG GTT CCG-3′. The annealing temperature was 58°C and the PCR was done using standard condition. The bi-cistronic operons expressing (*afR*/*afx*)/(*pdR*/*pdx*) and *cotB3* or *cotB4* were introduced into pACYC-Duet-1 using standard cloning techniques. Point mutations in Pdx protein from the redox system were introduced by following the QuickChange site-directed mutagenesis protocol. Primers for pdx_{D38E} mutation were: 5′-GAT ATT GTT GGT GAA TGT GGT GGT AGC G-3′ and 5′-CGC TAC CAC CAC ATT CAC CAA CAA TAT C-3′. Primers for the pdx_{W106E} mutation were: 5′-GAT GTT CCG GAT CGT CAG GAA TAA GCG GCC GCA TAA TG-3′ and 5′-CAT TAT GCG GCC GCT TAT TCC TGA CGA TCC GGA ACA TC-3′. The different pACYC-Duet-1 (*afR*/*afx* or *pdR*/*pdx*, *cotB3* or *cotB4*) plasmids were transformed into Bl21 (DE3) cells using standard protocols.

In vivo Assay for Comparison of Redox-Systems

The *in vivo* assay was used to determine the catalytic efficiency of CotB3 and CotB4 using different redox variants as shown in table 11. For the assay *E.coli* Bl21(DE3) cells harboring the different pACYC-Duet-1 variants were used. The metabolic precursors cyclooctat-9-en-7-ol (for pACYC-Duet-1 harboring CotB3) or cyclooct-9-en-5,7-diol (for pACYC-Duet-1 harboring CotB4) were added to the cultivation medium. After a 48 hours incubation time the amount of hydroxylated substrate was determined. Consequently, cyclooct-9-en-5,7-diol was determined to analyse the efficiency of CotB3. Alternativly, cyclooctatin was determined to evaluate the efficiency of CotB4.

Biomass for the assay was prepared by standard protein expression. From an overnight culture (LB medium, 34 µg/mL chloramphenicol) Bl21(DE3) cells harboring different pACYC-Duet-1 plasmid variants were grown in 100 mL baffled shake flask cultures at 37°C using 50 mL LB medium supplemented with 34 µg/mL chloramphenicol. As a negative control *E.Coli* Bl21 (DE3) harboring the empty pACYC-Duet-1 vector was carried out analogous. All strains were cultivated five times. After the cultures reached the OD_{600} 0.8, 1 mM δ-aminolevulinic acid, 1 mM $FeSO_4$ x $7H_2O$ and 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) was added. Protein expression was done for 24 hours at 30°C.

The *in vivo* assay was executed as follows. The cell density of the cultures was adjusted to OD_{600} 6 by centrifugation (3000 g for 5 minutes) and re-suspension in the appropriate volume. 10 mL of each culture (OD_{600} 6) was transferred to 50 mL baffled shake flasks and supplemented with 100 µM of cyclooctat-9-en-7-ol (for pACYC-Duet-1 harboring CotB3) or cyclooct-9-en-5,7-diol (for pACYC-Duet-1 harboring CotB4). Terpenes were added in acetonitrile to a final concentration of 0.1% (*v/v*). The cultures were sealed by membranes (Greiner bio-one Breathseal 676051) to avoid evaporation and incubated at 30°C for 2 days. To analyse the whole cell proteins 100 µL samples were taken every 6 hours. Samples for whole cell protein analysis were immediately centrifuged (10000 g for 5 minutes). The supernatant was discarded and the cell pellets were stored at -20°C. The samples were analyzed by a 12% SDS-Page gel according to Lämmili.

After the two day cultivation period the cultures were directly lysed in the culture medium by sonification on a Sonoplus HD2070 (Bandelin Electronic, Germany) performing 5 repeats on ice (5 min on and 3 min off at 80% power). The lysate was extracted 3 times with 5 mL ethyl acetate. All organic phases were combined and dried by $MgSO_4$. In the next step, the solvent was

removed under vacuum. The crude extract was solved in 1 mL ethyl acetate and analyzed by GC-MS. The GC-MS spectra were normalized to the total icon count of the internal standard α-humulene (Sigma-Aldrich, Germany). Quantification was done by determining the total ion count of cyclooctatin and cyclooct-9-en-5,7-diol, respectively.

Bioinformatics

***In Silico* Structure Prediction of CotB2**

The HHpred server and the Yasara bioinformatics suite was used to model the CotB2 structure. Modelling was conducted according to manufacturer's protocols and official guidelines. The substrate analog 2F-GGDP (2-fluoro- geranylgeranyldiphosphate) from the crystal structure of the taxadiene (PDB: 4OMG) was modelled into the *in silico* generated structure of CotB2 by using the MUSTANG algorithm of the Yasara bioinformatics Suite to align the α-domain of structure of taxadiene with CotB2.

Molecular Docking of Geranylgeranyl Diphosphate (GGPP) in the Crystal Structure of CotB2

Molecular Docking was performed using the AutoDock Vina program in combination with the AutoDockTools. A monomer of the crystal structure PDB: 4OMG was used as macromolecule. The structure of GGPP used for docking was energy optimized using MM2 force field ChemDraw Ultra Ver. 12.0 (CambridgeSoft). GGPP was docked into a simulation cell (Size: X-size = 30Å, Y-size = 30 Å, Z-size= 30 Å, angles: alpha = 90°, beta = 90°, gamma = 90°) around the center of the macromolecule (Center Grid Box: X center: -9.849, y center: 15.283, z center: -41.125). GGPP was docked with a total of 17 rotational bonds. Docking was done using standard parameters, except for "exhaustiveness", which was set to 100. Cluster analysis was done using the software pymol. The binding mode displaying the highest binding energy (Table 12) was used for modelling.

Table 12. Binding modes from AutoDock Vina program.

mode	Affinity/ kcal/mol)	dist from best mode r.m.s.d l.b	dist from best mode r.m.s.d u.b.
1	-8.1	0.000	0.000
2	-8.1	1.765	3.904
3	-8.0	1.822	4.178
4	-7.9	1.584	4.247
5	-7.9	1.064	1.752
6	-7.9	1.233	2.798

6. List of figures

7. List of tables

8. Appendix

Figure 1. (a) GC-MS chromatograms of fames derived from *T. oleaginosus* transformed with the plasmid pRF-HU2(GPD)-Fm1 (YPD medium, 72 hours cultivation). (**b**) Mass spectrum of α-linolenic acid methyl ester (RT 39.93 min)

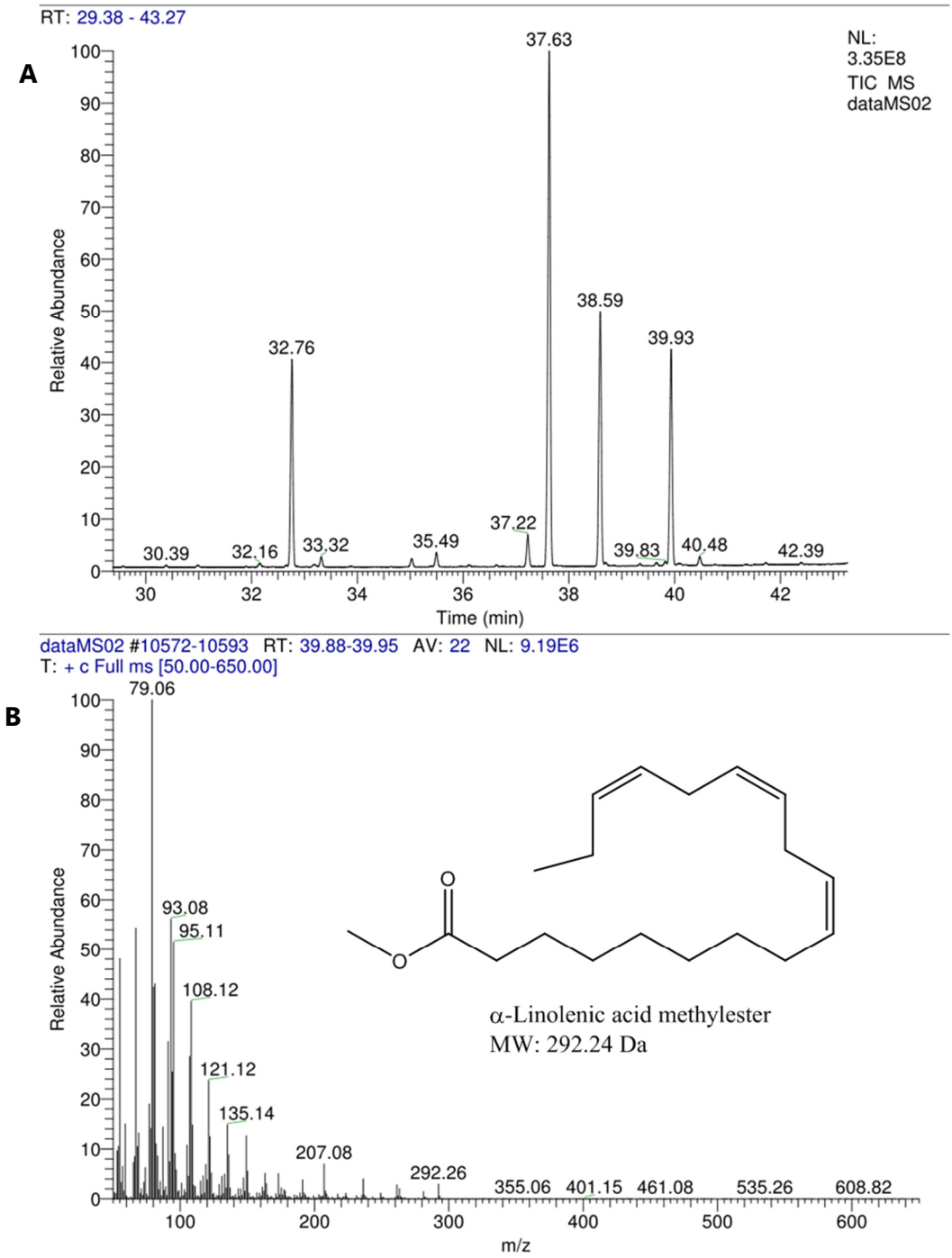


Figure 2. (**a**) GC-MS chromatograms of fames derived from *T. oleaginosus* transformed with the plasmid pRF-HU2(GPD)-IgASE2 (YPD medium, 72 hours cultivation) (**b**) Mass spectrum of eicosadienoic acid methyl ester (RT 42.67 min)

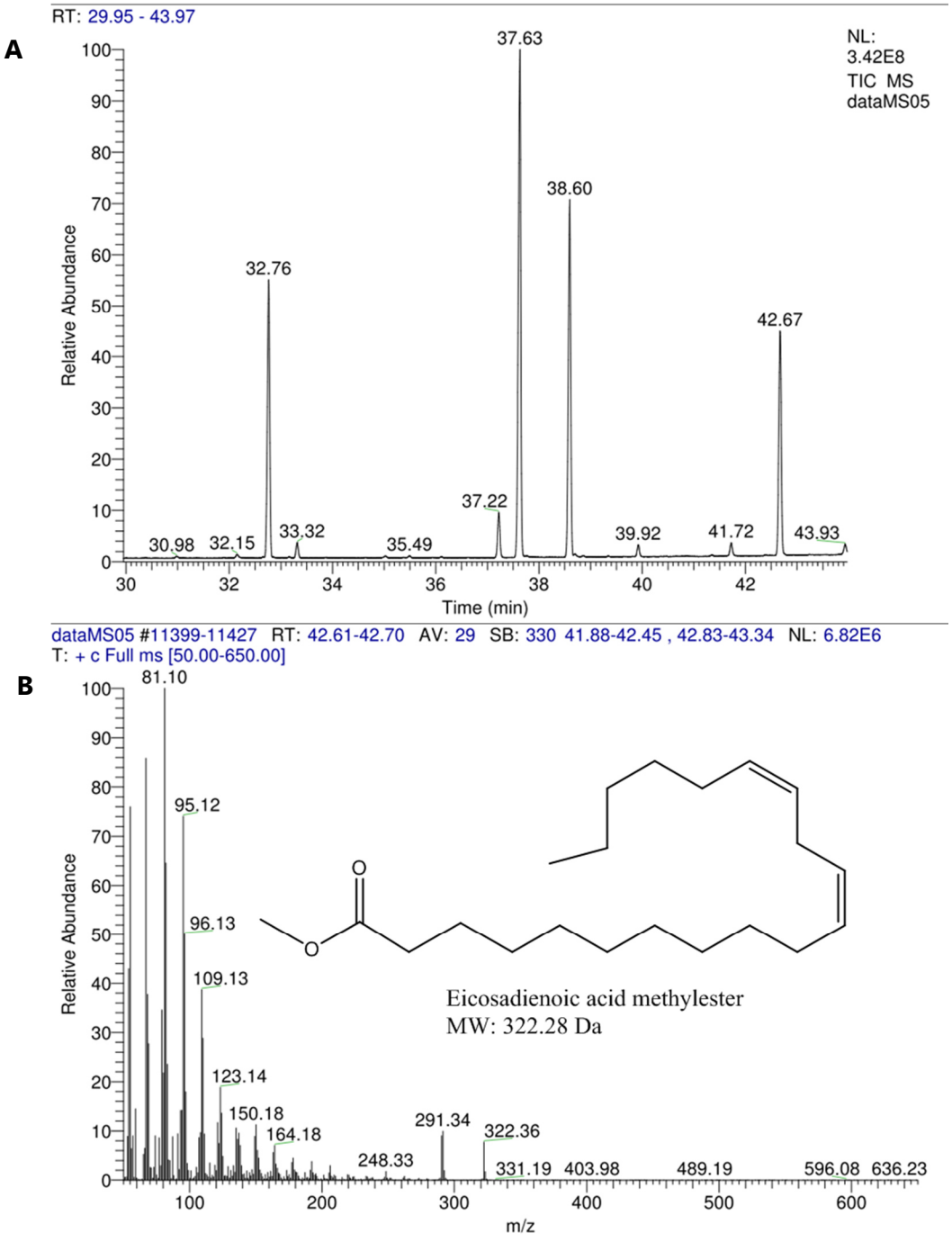


Figure 3. (**a**) GC-MS chromatograms of fames derived from *T. oleaginosus* strain II transformed with the plasmid pRF-HU2(GPD)-Fm1-IgASE2 (YPD medium, 72 hours cultivation) (**b**) Mass spectrum of eicosatrienoic acid methyl ester (RT 43.94 min)

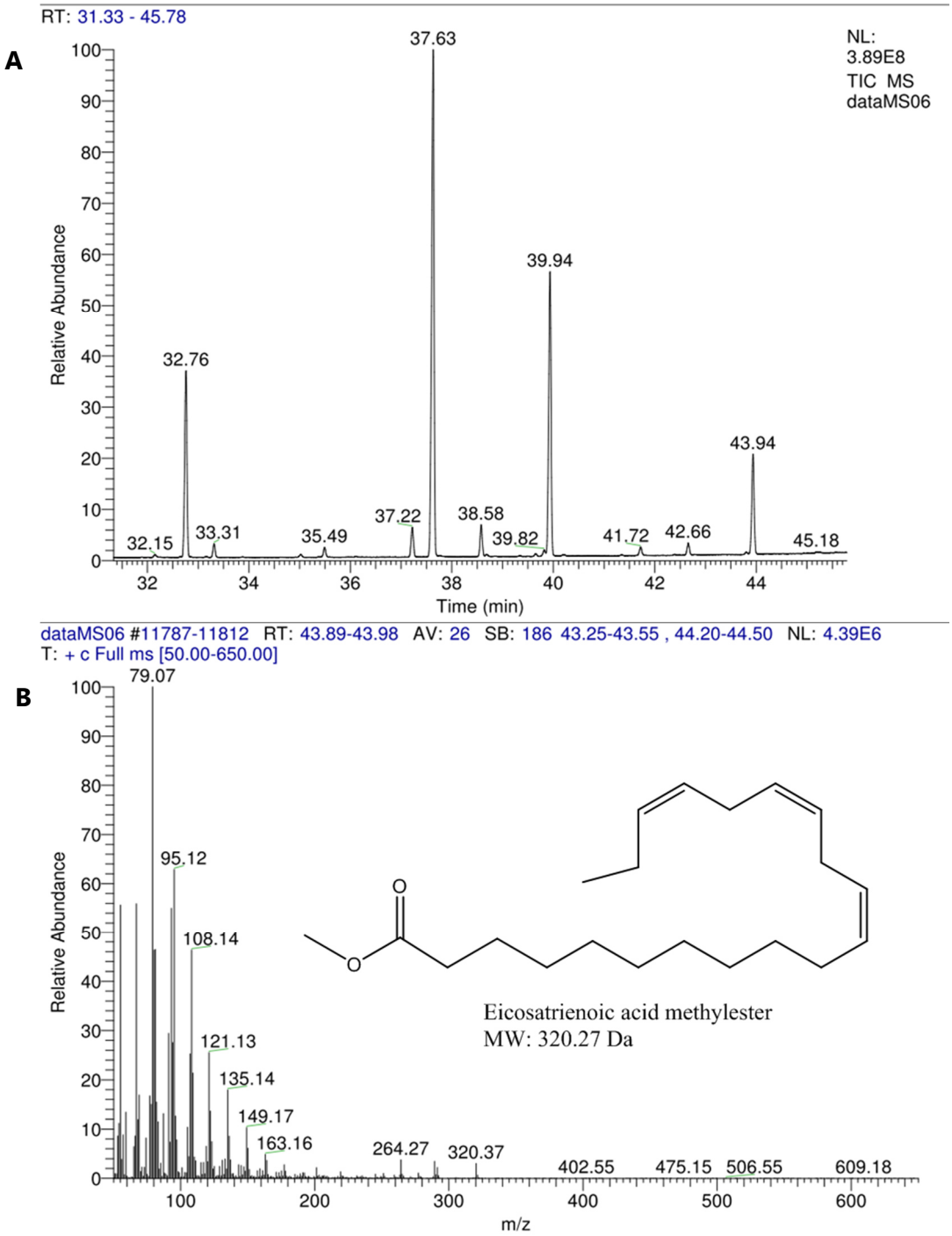


Figure 4. (**a**) GC-MS chromatograms of fames derived from *T. oleaginosus* transformed with the plasmid pRF-HU2(GPD)-PAI (YPD medium, 72 hours cultivation) (**b**) Mass spectrum of *E*10, *Z*12-linoleic acid methyl ester (RT 40.69 min)

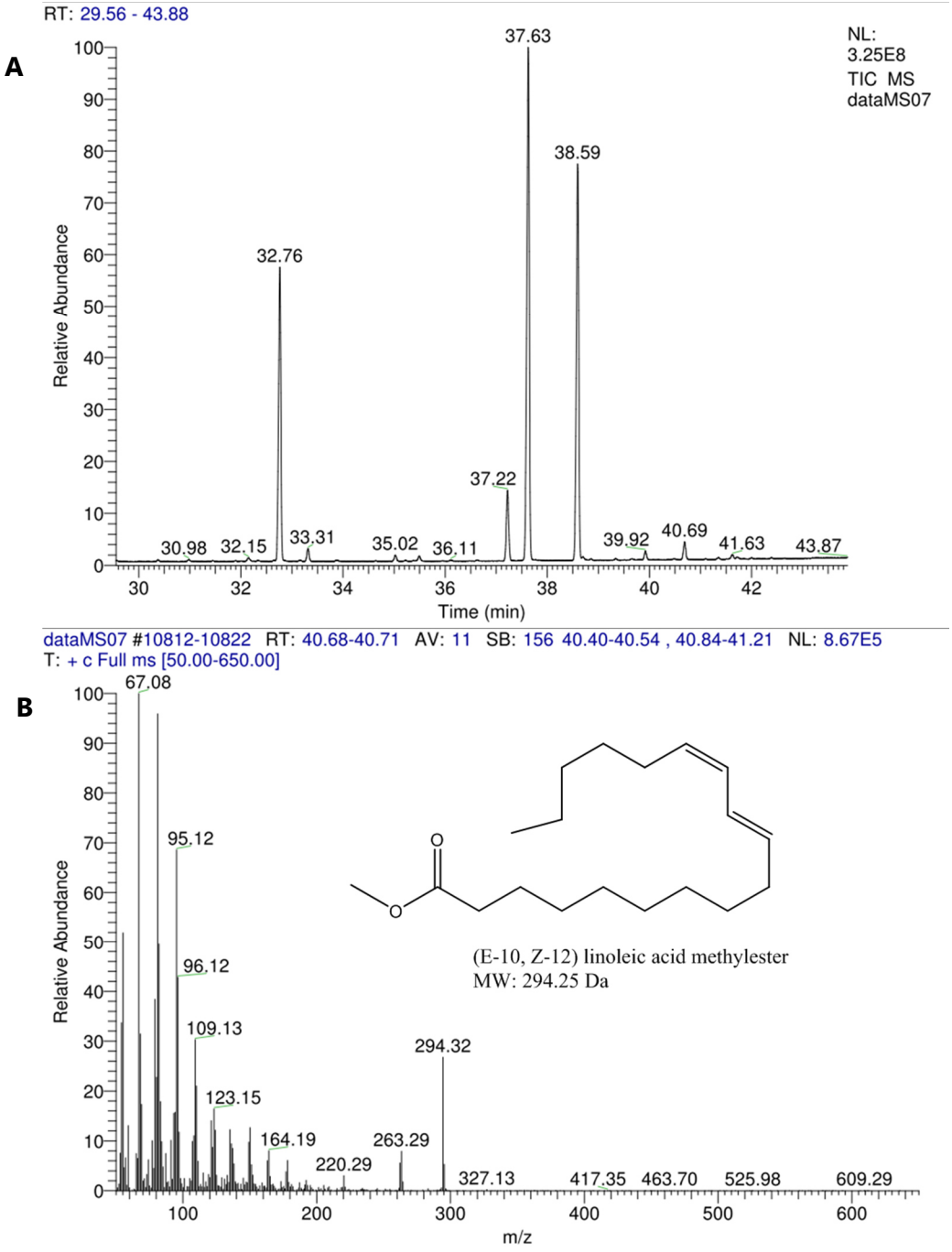


Table 1. Time depended fatty acid distribution in different media of wild type *T. oleaginosus* and *T. oleaginosus* expressing the Δ9 elongase IgASE2 from *I. galbana*, Δ12 /ω3 desaturase Fm1 from *F. moniliforme* and linoleic acid isomerase PAI from *P. acnes* and strain Ⅰ / Ⅱ of *T. oleaginosus* simultaneously expressing IgASE2 and Fm1. The standard deviation (SD [%]) is derived from three independent shake flask cultivations. (LP = Lipid production medium with either glucose, N-acetylglucosamine or xylose as carbon source.)

Media	Time [h]	Fatty acid	Wild type [%]	SD [%]	Fm1 [%]	SD [%]	IgASE2 [%]	SD [%]	Fm1 & IgASE2 strain I [%]	SD [%]	Fm1 & IgASE2 strain II [%]	SD [%]	PAI [%]	SD [%]
YPD medium	24	C16:0	17.23	2.30	16.24	0.44	18.87	1.42	17.86	0.81	14.34	0.20	22.28	0.33
		C18:0	3.66	1.33	3.04	0.29	4.07	0.77	2.59	0.15	2.20	0.05	5.67	0.08
		C18:1	37.46	2.99	40.51	0.80	35.31	3.19	37.84	0.87	45.68	0.37	35.24	0.71
		C18:2	37.66	5.41	25.17	0.25	27.48	2.82	16.68	0.80	2.51	1.94	32.57	0.92
		C18:3 (ALA)	4.00	1.16	15.04	1.78	2.64	0.64	13.16	1.34	29.48	2.14	2.58	0.09
		C20:2 (EDA)	0.00	0.00	0.00	0.00	10.17	1.54	7.51	0.42	0.93	0.28	0.00	0.00
		C20:3 (ETE)	0.00	0.00	0.00	0.00	1.46	0.31	4.35	0.83	4.85	0.22	0.00	0.00
		CLA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.66	0.26
	72	C16:0	13.52	0.74	14.50	1.03	18.10	0.50	15.05	0.50	14.92	0.20	17.56	0.89
		C18:0	1.71	0.56	2.33	0.04	2.70	0.17	1.63	0.09	2.27	0.06	3.89	0.38
		C18:1	35.69	0.23	42.45	0.50	32.34	2.81	36.33	0.79	42.29	1.02	35.11	0.55
		C18:2	46.25	1.79	19.66	1.31	27.66	1.99	11.32	1.16	2.21	0.39	39.55	1.16
		C18:3 (ALA)	2.83	0.49	21.06	2.00	1.34	0.44	16.98	1.65	28.46	1.46	1.33	0.04
		C20:2 (EDA)	0.00	0.00	0.00	0.00	16.81	0.59	9.73	0.43	0.85	0.12	0.00	0.00
		C20:3 (ETE)	0.00	0.00	0.00	0.00	1.04	0.27	8.95	1.07	9.00	0.27	0.00	0.00
		CLA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.56	0.13
	168	C16:0	10.56	0.71	11.47	0.54	13.65	1.41	11.87	0.17	13.20	0.06	13.89	0.12
		C18:0	1.36	0.36	1.49	0.19	1.44	0.42	0.97	0.09	1.59	0.06	2.35	0.17
		C18:1	39.46	0.84	51.84	1.26	34.32	2.92	43.92	0.07	48.66	0.52	39.80	0.62
		C18:2	46.63	0.78	17.55	1.35	31.64	3.15	10.82	0.51	1.51	0.08	41.34	0.80
		C18:3 (ALA)	1.99	0.56	17.64	0.82	0.99	0.38	13.21	0.40	25.11	0.26	0.94	0.06
		C20:2 (EDA)	0.00	0.00	0.00	0.00	16.93	1.90	10.01	0.28	0.80	0.12	0.00	0.00
		C20:3 (ETE)	0.00	0.00	0.00	0.00	1.03	0.29	9.20	0.58	9.13	0.18	0.00	0.00
		CLA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.68	0.28
LP on glucose	24	C16:0	27.10	0.96	26.30	0.10	21.89	1.88	29.07	0.25	28.23	0.69	28.30	0.62
		C18:0	15.00	1.77	15.43	2.04	20.32	1.66	14.97	2.09	19.72	4.38	13.51	0.29
		C18:1	48.10	0.76	49.75	2.01	48.59	0.22	45.09	2.59	42.61	3.60	46.12	0.23
		C18:2	8.62	0.19	3.44	0.48	5.33	0.18	2.30	0.30	0.71	0.03	10.29	0.15
		C18:3 (ALA)	1.17	0.13	5.08	0.56	0.60	0.01	5.31	0.21	7.72	0.10	1.44	0.03
		C20:2 (EDA)	0.00	0.00	0.00	0.00	2.98	0.04	1.42	0.22	0.25	0.16	0.00	0.00
		C20:3 (ETE)	0.00	0.00	0.00	0.00	0.28	0.01	1.85	0.02	0.77	0.08	0.00	0.00
		CLA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.35	0.04
	72	C16:0	29.03	1.06	29.04	0.81	24.06	1.62	31.99	0.00	28.75	0.37	31.27	0.17
		C18:0	13.84	3.09	13.64	2.29	15.80	1.44	12.46	2.94	25.54	1.88	9.70	0.15
		C18:1	50.22	2.21	51.06	1.82	50.36	0.52	47.22	2.35	39.42	1.22	49.36	0.02
		C18:2	6.20	0.47	3.08	0.26	5.37	0.10	1.71	0.05	0.54	0.03	8.55	0.02
		C18:3 (ALA)	0.70	0.09	3.19	0.23	0.45	0.04	3.94	0.67	4.95	0.31	0.95	0.04
		C20:2 (EDA)	0.00	0.00	0.00	0.00	3.68	0.12	1.17	0.13	0.11	0.00	0.00	0.00
		C20:3 (ETE)	0.00	0.00	0.00	0.00	0.29	0.02	1.52	0.09	0.69	0.05	0.00	0
		CLA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.17	0.01

	168	C16:0	29.02	0.56	29.33	0.44	23.61	1.53	31.39	0.08	28.45	0.32	30.81	0.23
		C18:0	13.31	3.01	12.32	1.95	15.32	1.33	12.14	2.68	23.77	1.22	9.44	0.17
		C18:1	50.05	2.08	51.17	2.25	50.05	0.49	47.56	2.35	40.42	0.70	49.56	0.03
		C18:2	6.94	0.47	4.47	2.48	5.73	0.08	1.68	0.06	0.53	0.05	9.03	0.02
		C18:3 (ALA)	0.69	0.11	2.71	1.72	0.42	0.03	4.10	0.50	5.78	0.21	0.92	0.04
		C20:2 (EDA)	0.00	0.00	0.00	0.00	4.55	0.12	1.27	0.15	0.12	0.00	0.00	0.00
		C20:3 (ETE)	0.00	0.00	0.00	0.00	0.33	0.01	1.86	0.11	0.92	0.07	0.00	0.00
		CLA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.23	0.01
LP on N-acetylglucosamine	24	C16:0	21.86	1.25	23.68	1.50	28.62	0.24	23.94	2.49	24.27	1.05	27.28	1.42
		C18:0	21.82	1.22	21.89	1.06	15.53	0.37	19.34	2.05	26.56	1.44	17.52	0.73
		C18:1	49.60	0.52	48.31	0.83	48.00	0.30	48.89	0.92	41.71	0.31	44.67	1.00
		C18:2	6.02	0.38	2.73	0.32	5.06	0.21	1.75	0.21	0.60	0.00	8.93	0.35
		C18:3 (ALA)	0.69	0.09	3.39	0.15	0.55	0.05	4.04	0.19	5.96	0.17	1.04	0.02
		C20:2 (EDA)	0.00	0.00	0.00	0.00	2.02	0.19	0.76	0.05	0.30	0.26	0.00	0.00
		C20:3 (ETE)	0.00	0.00	0.00	0.00	0.22	0.03	1.28	0.04	0.60	0.02	0.00	0.00
		CLA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.56	0.07
	72	C16:0	25.21	0.60	25.84	1.48	31.40	1.47	28.48	1.66	30.35	2.55	29.86	0.05
		C18:0	16.48	0.25	16.61	1.14	14.70	0.83	17.64	1.44	23.52	1.61	16.57	0.32
		C18:1	52.00	0.66	51.51	1.22	46.21	2.14	47.07	0.99	39.50	1.30	44.91	0.25
		C18:2	5.68	0.17	2.50	0.39	4.66	0.32	1.22	0.28	0.46	0.08	7.79	0.11
		C18:3 (ALA)	0.64	0.01	3.55	0.06	0.48	0.03	3.19	0.28	5.46	0.33	0.78	0.00
		C20:2 (EDA)	0.00	0.00	0.00	0.00	2.31	0.18	0.76	0.12	0.10	0.00	0.00	0.00
		C20:3 (ETE)	0.00	0.00	0.00	0.00	0.23	0.02	1.64	0.06	0.61	0.05	0.00	0.00
		CLA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.00
	168	C16:0	26.18	1.22	25.83	1.80	30.50	0.28	27.18	1.84	27.86	0.69	29.14	1.06
		C18:0	15.93	0.47	15.73	1.19	14.16	0.63	16.82	1.20	22.68	0.86	15.42	0.45
		C18:1	51.07	1.93	52.16	1.38	47.49	0.55	48.99	1.34	42.53	0.04	46.18	1.10
		C18:2	6.09	0.28	2.59	0.42	4.81	0.39	1.25	0.28	0.46	0.05	8.35	0.45
		C18:3 (ALA)	0.73	0.03	3.70	0.09	0.49	0.04	3.31	0.25	5.73	0.18	0.82	0.05
		C20:2 (EDA)	0.00	0.00	0.00	0.00	2.32	0.19	0.78	0.12	0.10	0.00	0.00	0.00
		C20:3 (ETE)	0.00	0.00	0.00	0.00	0.24	0.03	1.67	0.06	0.64	0.02	0.00	0.00
		CLA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.09	0.01
LP on xylose	24	C16:0	27.37	0.60	26.15	1.45	28.40	0.59	29.34	0.64	23.73	1.48	27.21	1.60
		C18:0	16.05	0.52	17.41	2.51	18.20	0.81	15.68	0.70	27.48	1.84	16.65	0.92
		C18:1	40.54	0.30	44.00	1.16	40.37	1.51	41.52	1.17	37.38	0.23	42.26	0.81
		C18:2	14.21	0.24	5.33	0.34	7.82	0.23	3.21	0.09	0.94	0.05	11.90	0.04
		C18:3 (ALA)	1.84	0.01	7.10	0.46	1.02	0.01	6.43	0.03	9.01	0.34	1.41	0.02
		C20:2 (EDA)	0.00	0.00	0.00	0.00	3.79	0.09	1.71	0.10	0.17	0.00	0.00	0.00
		C20:3 (ETE)	0.00	0.00	0.00	0.00	0.41	0.01	2.12	0.18	1.28	0.07	0.00	0.00
		CLA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.57	0.08
	72	C16:0	33.49	0.04	29.94	1.45	30.86	0.08	31.85	0.11	27.08	1.56	31.21	0.06
		C18:0	15.04	1.51	12.72	2.75	13.73	1.05	12.03	0.70	24.24	0.43	11.30	0.49
		C18:1	43.48	1.13	47.71	1.19	45.70	0.97	46.63	0.72	39.12	2.15	46.82	0.34
		C18:2	7.21	0.35	3.98	0.23	5.42	0.13	1.97	0.10	0.50	0.05	9.50	0.18
		C18:3 (ALA)	0.77	0.03	5.64	0.63	0.55	0.04	4.42	0.30	5.30	0.60	0.93	0.01
		C20:2 (EDA)	0.00	0.00	0.00	0.00	3.43	0.05	1.44	0.17	0.12	0.01	0.00	0.00
		C20:3 (ETE)	0.00	0.00	0.00	0.00	0.32	0.00	1.66	0.32	0.91	0.05	0.00	0.00

		CLA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.23	0.01
	168	**C16:0**	32.79	0.10	29.29	1.36	30.01	0.31	31.12	0.10	27.44	0.30	30.49	0.19
		C18:0	13.99	1.31	12.36	2.61	13.07	0.87	11.64	0.68	24.00	0.68	11.23	0.29
		C18:1	44.17	0.97	48.31	1.12	45.95	0.83	46.96	0.55	40.77	0.24	47.44	0.56
		C18:2	8.27	0.23	4.05	0.22	5.81	0.12	1.93	0.08	0.52	0.03	9.70	0.16
		C18:3 (ALA)	0.78	0.03	6.00	0.67	0.53	0.04	4.62	0.35	6.01	0.14	0.86	0.03
		C20:2 (EDA)	0.00	0.00	0.00	0.00	4.25	0.10	1.55	0.18	0.13	0.01	0.00	0.00
		C20:3 (ETE)	0.00	0.00	0.00	0.00	0.37	0.01	2.18	0.07	1.13	0.03	0.00	0.00
		CLA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.28	0.01

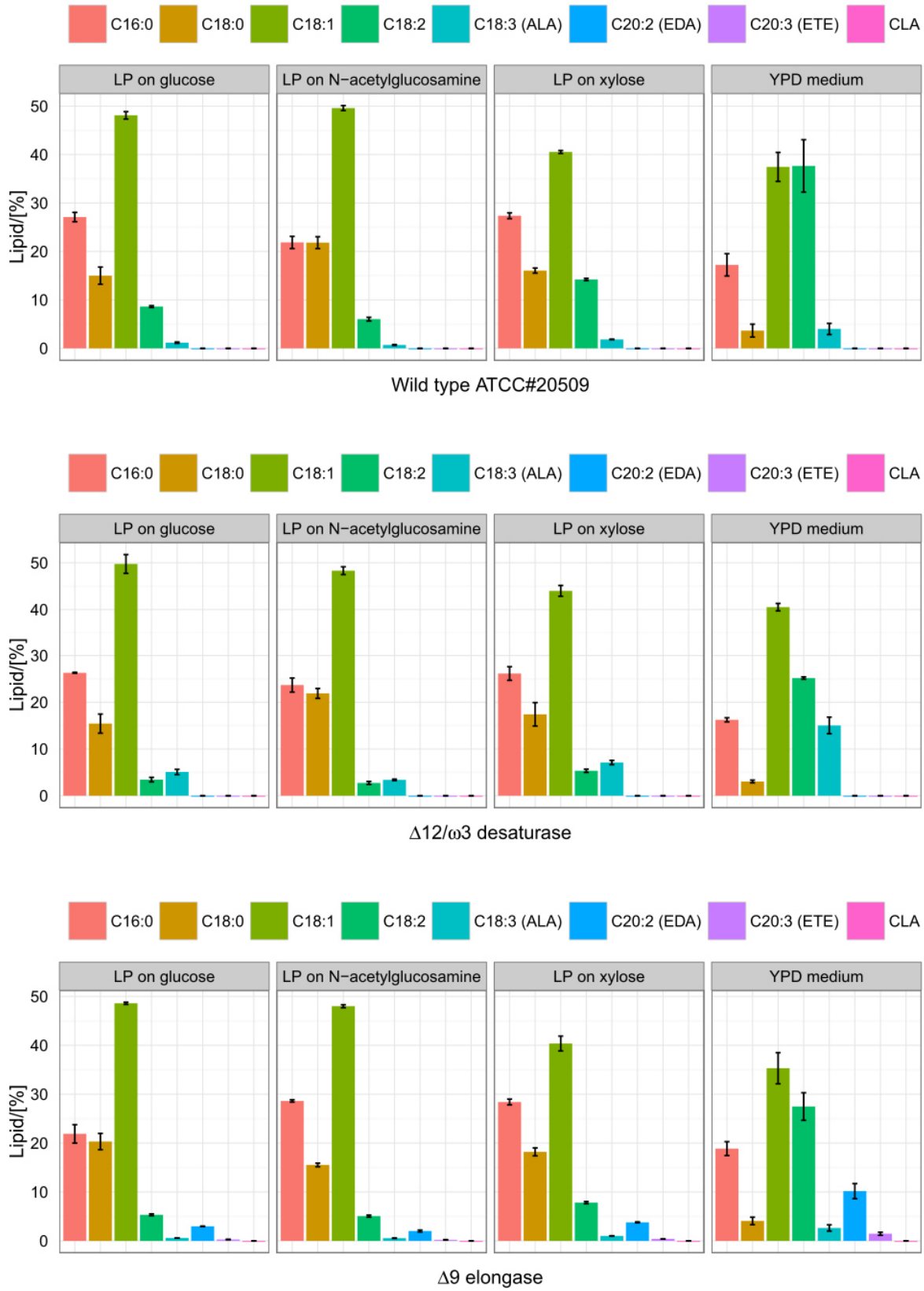
Table 2. Total fatty acid content referred to the total lipid content per DCW. Fatty acid distribution as total values (ABS) in different media of wild type *T. oleaginosus* and *T. oleaginosus* expressing the Δ9 elongase IgASE2 from *I. galbana*, Δ12 /ω3 desaturase Fm1 from *F. moniliforme* and linoleic acid isomerase PAI from *P. acnes* and strain I / II of *T. oleaginosus* simultaneously expressing IgASE2 and Fm1. The standard deviation (SD ABS [%]) is derived from three independent shake flask cultivations. Total lipid content was dependent on the media type, carbon source and the cultivation time peaking after 72 h (YPD medium: 10% DCW, LP on glucose: 48% DCW, LP on N-acetylglucosamine: 35% DCW, LP on xylose: 57% DCW). About half of the total lipid yield was obtained after 24 h. With an average biomass formation of 13 g/L (DCW) values of Table S2 allow for calculating the absolute yield of each individual fatty acid. (LP = Lipid production medium with either glucose, N-acetylglucosamine or xylose as carbon source.)

Media	Time [h]	Fatty acid	Wild type ABS [%]	SD ABS [%]	Fm1 ABS [%]	SD ABS [%]	IgASE2 ABS [%]	SD ABS [%]	Fm1 & IgASE2 strain I ABS [%]	SD ABS [%]	Fm1 & IgASE2 strain II ABS [%]	SD ABS [%]	PAI ABS [%]	SD ABS [%]
YPD medium	24	C16:0	0.86	0.12	0.81	0.02	0.94	0.07	0.89	0.04	0.72	0.01	1.11	0.02
		C18:0	0.18	0.07	0.15	0.01	0.20	0.04	0.13	0.01	0.11	0.00	0.28	0.00
		C18:1	1.87	0.15	2.03	0.04	1.77	0.16	1.89	0.04	2.28	0.02	1.76	0.04
		C18:2	1.88	0.27	1.26	0.01	1.37	0.14	0.83	0.04	0.13	0.10	1.63	0.05
		C18:3 (ALA)	0.20	0.06	0.75	0.09	0.13	0.03	0.66	0.07	1.47	0.11	0.13	0.00
		C20:2 (EDA)	0.00	0.00	0.00	0.00	0.51	0.08	0.38	0.02	0.05	0.01	0.00	0.00
		C20:3 (ETE)	0.00	0.00	0.00	0.00	0.07	0.02	0.22	0.04	0.24	0.01	0.00	0.00
		CLA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.01
	72	C16:0	1.35	0.07	1.45	0.10	1.81	0.05	1.51	0.05	1.49	0.02	1.76	0.09
		C18:0	0.17	0.06	0.23	0.00	0.27	0.02	0.16	0.01	0.23	0.01	0.39	0.04
		C18:1	3.57	0.02	4.25	0.05	3.23	0.28	3.63	0.08	4.23	0.10	3.51	0.06
		C18:2	4.63	0.18	1.97	0.13	2.77	0.20	1.13	0.12	0.22	0.04	3.96	0.12
		C18:3 (ALA)	0.28	0.05	2.11	0.20	0.13	0.04	1.70	0.17	2.85	0.15	0.13	0.00
		C20:2 (EDA)	0.00	0.00	0.00	0.00	1.68	0.06	0.97	0.04	0.09	0.01	0.00	0.00
		C20:3 (ETE)	0.00	0.00	0.00	0.00	0.10	0.03	0.90	0.11	0.90	0.03	0.00	0.00
		CLA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.26	0.01
	168	C16:0	1.06	0.07	1.15	0.05	1.37	0.14	1.19	0.02	1.32	0.01	1.39	0.01
		C18:0	0.14	0.04	0.15	0.02	0.14	0.04	0.10	0.01	0.16	0.01	0.24	0.02
		C18:1	3.95	0.08	5.18	0.13	3.43	0.29	4.39	0.01	4.87	0.05	3.98	0.06
		C18:2	4.66	0.08	1.76	0.14	3.16	0.32	1.08	0.05	0.15	0.01	4.13	0.08
		C18:3 (ALA)	0.20	0.06	1.76	0.08	0.10	0.04	1.32	0.04	2.51	0.03	0.09	0.01
		C20:2 (EDA)	0.00	0.00	0.00	0.00	1.69	0.19	1.00	0.03	0.08	0.01	0.00	0.00
		C20:3 (ETE)	0.00	0.00	0.00	0.00	0.10	0.03	0.92	0.06	0.91	0.02	0.00	0.00
		CLA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.17	0.03
LP on glucose	24	C16:0	6.50	0.23	6.31	0.02	5.25	0.45	6.98	0.06	6.77	0.17	6.79	0.15
		C18:0	3.60	0.42	3.70	0.49	4.88	0.40	3.59	0.50	4.73	1.05	3.24	0.07
		C18:1	11.54	0.18	11.94	0.48	11.66	0.05	10.82	0.62	10.23	0.86	11.07	0.06
		C18:2	2.07	0.05	0.83	0.12	1.28	0.04	0.55	0.07	0.17	0.01	2.47	0.04
		C18:3 (ALA)	0.28	0.03	1.22	0.13	0.14	0.00	1.27	0.05	1.85	0.02	0.35	0.01
		C20:2 (EDA)	0.00	0.00	0.00	0.00	0.72	0.01	0.34	0.05	0.06	0.04	0.00	0.00
		C20:3 (ETE)	0.00	0.00	0.00	0.00	0.07	0.00	0.44	0.00	0.18	0.02	0.00	0.00
		CLA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.01
	72	C16:0	13.93	0.51	13.94	0.39	11.55	0.78	15.36	0.00	13.80	0.18	15.01	0.08
		C18:0	6.64	1.48	6.55	1.10	7.58	0.69	5.98	1.41	12.26	0.90	4.66	0.07
		C18:1	24.11	1.06	24.51	0.87	24.17	0.25	22.67	1.13	18.92	0.59	23.69	0.01
		C18:2	2.98	0.23	1.48	0.12	2.58	0.05	0.82	0.02	0.26	0.01	4.10	0.01

Media	Time [h]	Fatty acid	Wild type ABS [%]	SD ABS [%]	Fm1 ABS [%]	SD ABS [%]	IgASE2 ABS [%]	SD ABS [%]	Fm1 & IgASE2 strain I ABS [%]	SD ABS [%]	Fm1 & IgASE2 strain II ABS [%]	SD ABS [%]	PAI ABS [%]	SD ABS [%]
LP on glucose		C18:3 (ALA)	0.34	0.04	1.53	0.11	0.22	0.02	1.89	0.32	2.38	0.15	0.46	0.02
		C20:2 (EDA)	0.00	0.00	0.00	0.00	1.77	0.06	0.56	0.06	0.05	0.00	0.00	0.00
		C20:3 (ETE)	0.00	0.00	0.00	0.00	0.14	0.01	0.73	0.04	0.33	0.02	0.00	0.00
		CLA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.00
	168	C16:0	13.93	0.27	14.08	0.21	11.33	0.73	15.07	0.04	13.66	0.15	14.79	0.11
		C18:0	6.39	1.44	5.91	0.94	7.35	0.64	5.83	1.29	11.41	0.59	4.53	0.08
		C18:1	24.02	1.00	24.56	1.08	24.02	0.24	22.83	1.13	19.40	0.34	23.79	0.01
		C18:2	3.33	0.23	2.15	1.19	2.75	0.04	0.81	0.03	0.25	0.02	4.33	0.01
		C18:3 (ALA)	0.33	0.05	1.30	0.83	0.20	0.01	1.97	0.24	2.77	0.10	0.44	0.02
		C20:2 (EDA)	0.00	0.00	0.00	0.00	2.18	0.06	0.61	0.07	0.06	0.00	0.00	0.00
		C20:3 (ETE)	0.00	0.00	0.00	0.00	0.16	0.00	0.89	0.05	0.44	0.03	0.00	0.00
		CLA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.11	0.00
LP on N-acetylglucosamine	24	C16:0	3.72	0.21	4.03	0.26	4.87	0.04	4.07	0.42	4.13	0.18	4.64	0.24
		C18:0	3.71	0.21	3.72	0.18	2.64	0.06	3.29	0.35	4.52	0.24	2.98	0.12
		C18:1	8.43	0.09	8.21	0.14	8.16	0.05	8.31	0.16	7.09	0.05	7.59	0.17
		C18:2	1.02	0.06	0.46	0.05	0.86	0.04	0.30	0.04	0.10	0.00	1.52	0.06
		C18:3 (ALA)	0.12	0.02	0.58	0.03	0.09	0.01	0.69	0.03	1.01	0.03	0.18	0.00
		C20:2 (EDA)	0.00	0.00	0.00	0.00	0.34	0.03	0.13	0.01	0.05	0.04	0.00	0.00
		C20:3 (ETE)	0.00	0.00	0.00	0.00	0.04	0.01	0.22	0.01	0.10	0.00	0.00	0.00
		CLA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.09	0.01
	72	C16:0	8.57	0.20	8.79	0.50	10.68	0.50	9.68	0.56	10.32	0.87	10.15	0.02
		C18:0	5.60	0.09	5.65	0.39	5.00	0.28	6.00	0.49	8.00	0.55	5.63	0.11
		C18:1	17.68	0.22	17.51	0.41	15.71	0.73	16.00	0.34	13.43	0.44	15.27	0.09
		C18:2	1.93	0.06	0.85	0.13	1.58	0.11	0.41	0.10	0.16	0.03	2.65	0.04
		C18:3 (ALA)	0.22	0.00	1.21	0.02	0.16	0.01	1.08	0.10	1.86	0.11	0.27	0.00
		C20:2 (EDA)	0.00	0.00	0.00	0.00	0.79	0.06	0.26	0.04	0.03	0.00	0.00	0.00
		C20:3 (ETE)	0.00	0.00	0.00	0.00	0.08	0.01	0.56	0.02	0.21	0.02	0.00	0.00
		CLA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00
	168	C16:0	8.90	0.41	8.78	0.61	10.37	0.10	9.24	0.63	9.47	0.23	9.91	0.36
		C18:0	5.42	0.16	5.35	0.40	4.81	0.21	5.72	0.41	7.71	0.29	5.24	0.15
		C18:1	17.36	0.66	17.73	0.47	16.15	0.19	16.66	0.46	14.46	0.01	15.70	0.37
		C18:2	2.07	0.10	0.88	0.14	1.64	0.13	0.43	0.10	0.16	0.02	2.84	0.15
		C18:3 (ALA)	0.25	0.01	1.26	0.03	0.17	0.01	1.13	0.09	1.95	0.06	0.28	0.02
		C20:2 (EDA)	0.00	0.00	0.00	0.00	0.79	0.06	0.27	0.04	0.03	0.00	0.00	0.00
		C20:3 (ETE)	0.00	0.00	0.00	0.00	0.08	0.01	0.57	0.02	0.22	0.01	0.00	0.00
		CLA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00
LP on xylose	24	C16:0	7.66	0.17	7.32	0.41	7.95	0.17	8.21	0.18	6.64	0.41	7.62	0.45
		C18:0	4.49	0.15	4.88	0.70	5.10	0.23	4.39	0.20	7.69	0.52	4.66	0.26
		C18:1	11.35	0.08	12.32	0.32	11.30	0.42	11.62	0.33	10.47	0.06	11.83	0.23
		C18:2	3.98	0.07	1.49	0.10	2.19	0.06	0.90	0.03	0.26	0.01	3.33	0.01
		C18:3 (ALA)	0.51	0.00	1.99	0.13	0.29	0.00	1.80	0.01	2.52	0.10	0.39	0.01
		C20:2 (EDA)	0.00	0.00	0.00	0.00	1.06	0.03	0.48	0.03	0.05	0.00	0.00	0.00
		C20:3 (ETE)	0.00	0.00	0.00	0.00	0.11	0.00	0.59	0.05	0.36	0.02	0.00	0.00
		CLA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.16	0.02
	72	C16:0	18.75	0.02	16.77	0.81	17.28	0.04	17.84	0.06	15.16	0.87	17.48	0.03

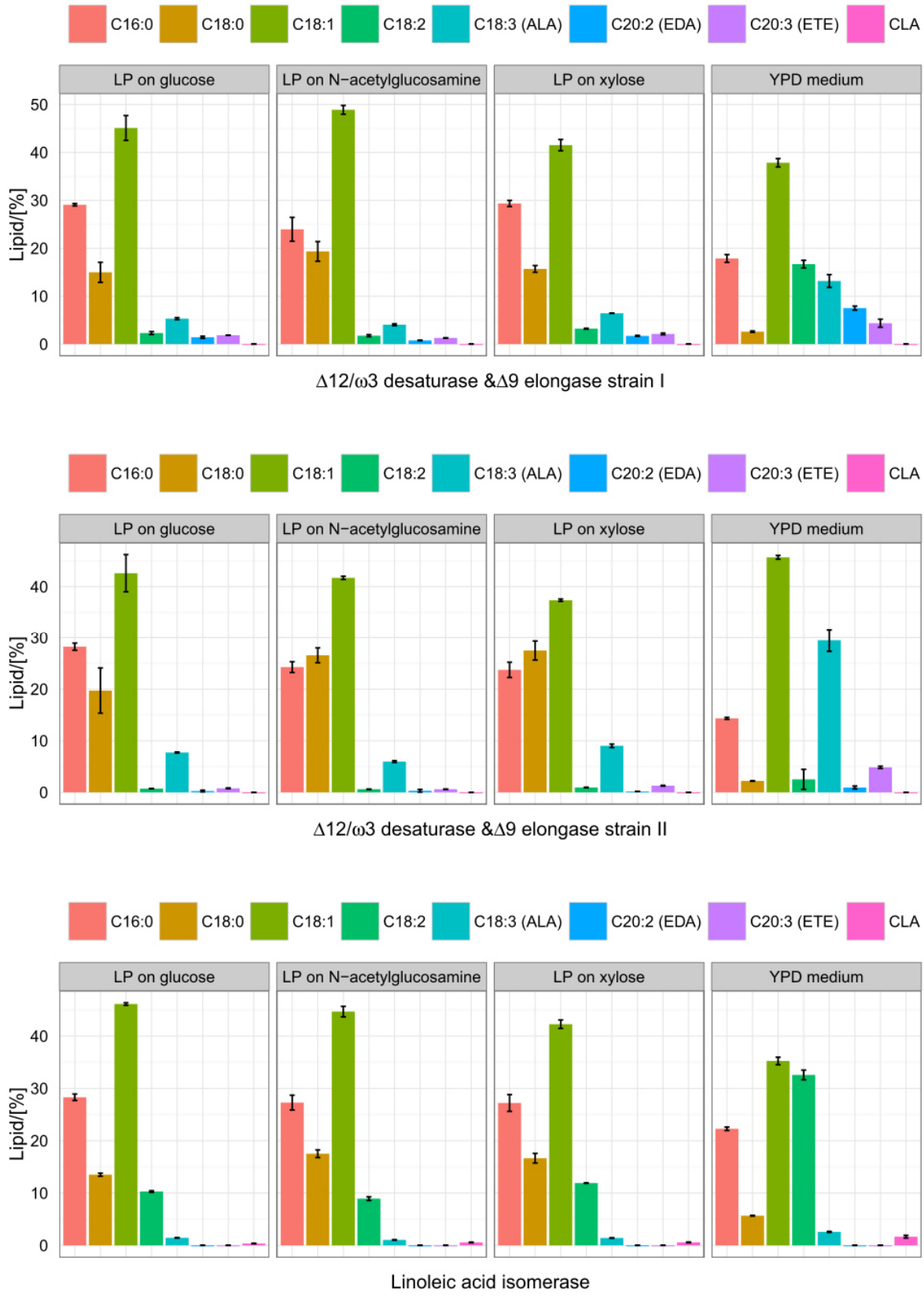
Media	Time [h]	Fatty acid	Wild type ABS [%]	SD ABS [%]	Fm1 ABS [%]	SD ABS [%]	IgASE2 ABS [%]	SD ABS [%]	Fm1 & IgASE2 strain I ABS [%]	SD ABS [%]	Fm1 & IgASE2 strain II ABS [%]	SD ABS [%]	PAI ABS [%]	SD ABS [%]
LP on xylose		C18:0	8.42	0.85	7.12	1.54	7.69	0.59	6.74	0.39	13.57	0.24	6.33	0.27
		C18:1	24.35	0.63	26.72	0.67	25.59	0.54	26.11	0.40	21.91	1.20	26.22	0.19
		C18:2	4.04	0.20	2.23	0.13	3.04	0.07	1.10	0.06	0.28	0.03	5.32	0.10
		C18:3 (ALA)	0.43	0.02	3.16	0.35	0.31	0.02	2.48	0.17	2.97	0.34	0.52	0.01
		C20:2 (EDA)	0.00	0.00	0.00	0.00	1.92	0.03	0.81	0.10	0.07	0.01	0.00	0.00
		C20:3 (ETE)	0.00	0.00	0.00	0.00	0.18	0.00	0.93	0.18	0.51	0.03	0.00	0.00
		CLA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.13	0.01
	168	C16:0	18.36	0.06	16.40	0.76	16.81	0.17	17.43	0.06	15.37	0.17	17.07	0.11
		C18:0	7.83	0.73	6.92	1.46	7.32	0.49	6.52	0.38	13.44	0.38	6.29	0.16
		C18:1	24.74	0.54	27.05	0.63	25.73	0.46	26.30	0.31	22.83	0.13	26.57	0.31
		C18:2	4.63	0.13	2.27	0.12	3.25	0.07	1.08	0.04	0.29	0.02	5.43	0.09
		C18:3 (ALA)	0.44	0.02	3.36	0.38	0.30	0.02	2.59	0.20	3.37	0.08	0.48	0.02
		C20:2 (EDA)	0.00	0.00	0.00	0.00	2.38	0.06	0.87	0.10	0.07	0.01	0.00	0.00
		C20:3 (ETE)	0.00	0.00	0.00	0.00	0.21	0.01	1.22	0.04	0.63	0.02	0.00	0.00
		CLA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.16	0.01

Figure 5. Graphical illustration of Table S1 of the fatty acid distribution after 24 hours cultivation



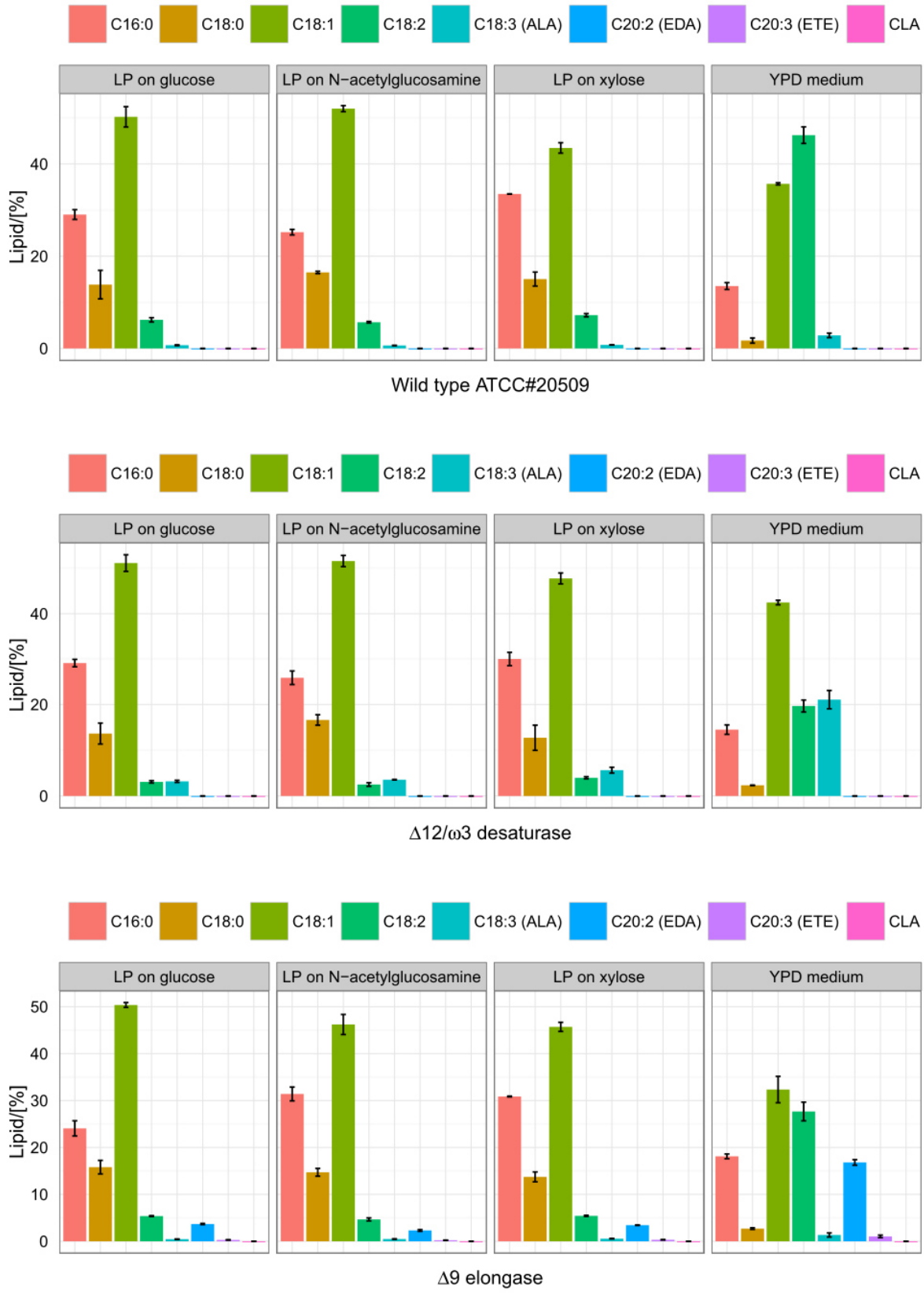
(LP = Lipid production medium with either glycose, N-acetylglucosamine or xylose as carbon source)

Figure 6. Graphical illustration of Table S1 of the fatty acid distribution after 24 hours cultivation.



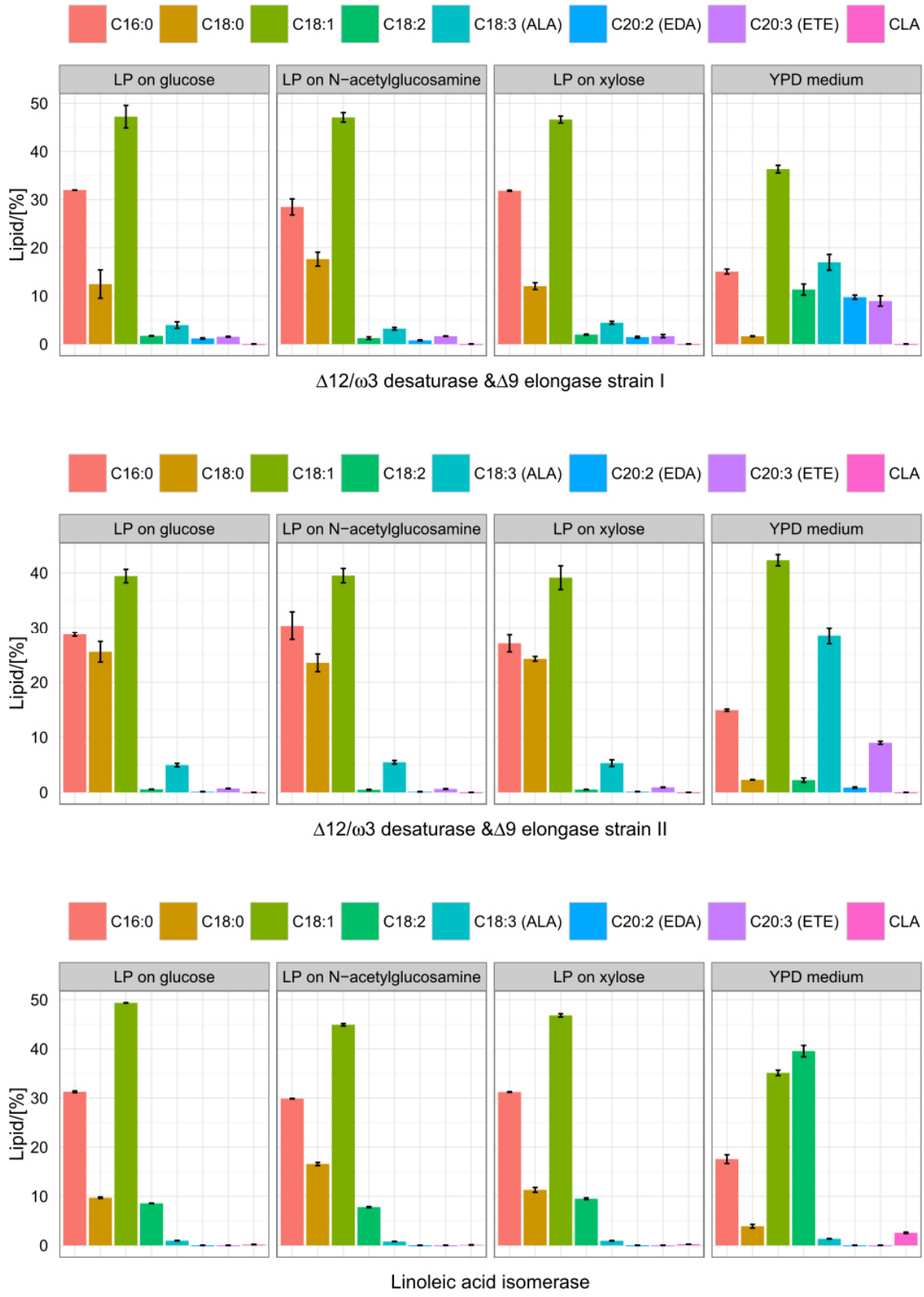
(LP = Lipid production medium with either glycose, N-acetylglucosamine or xylose as carbon source)

Figure 7. Graphical illustration of Table S1 of the fatty acid distribution after 72 hours cultivation.



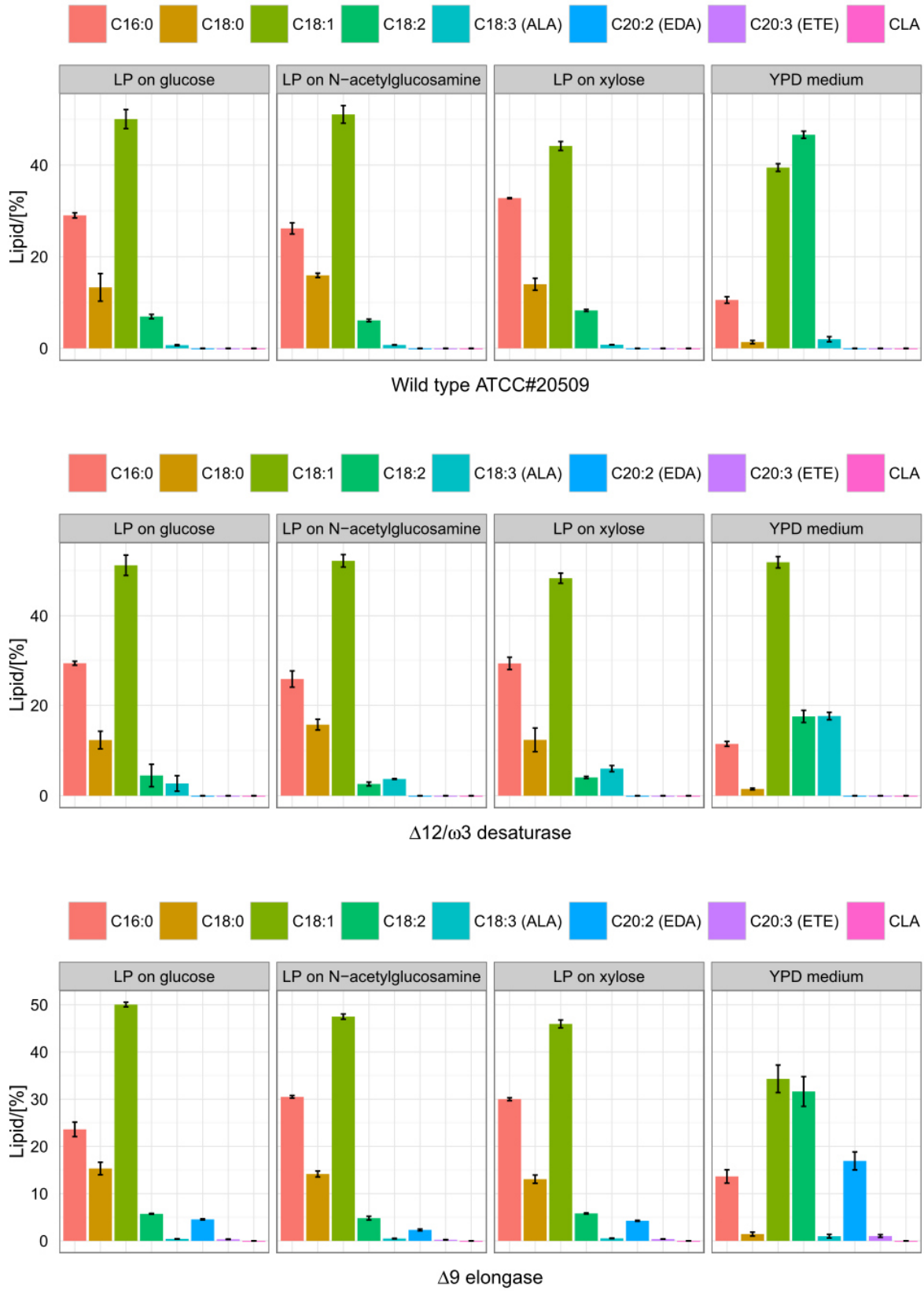
(LP = Lipid production medium with either glycose, N-acetylglucosamine or xylose as carbon source)

Figure 8. Graphical illustration of Table S1 of the fatty acid distribution after 72 hours cultivation.



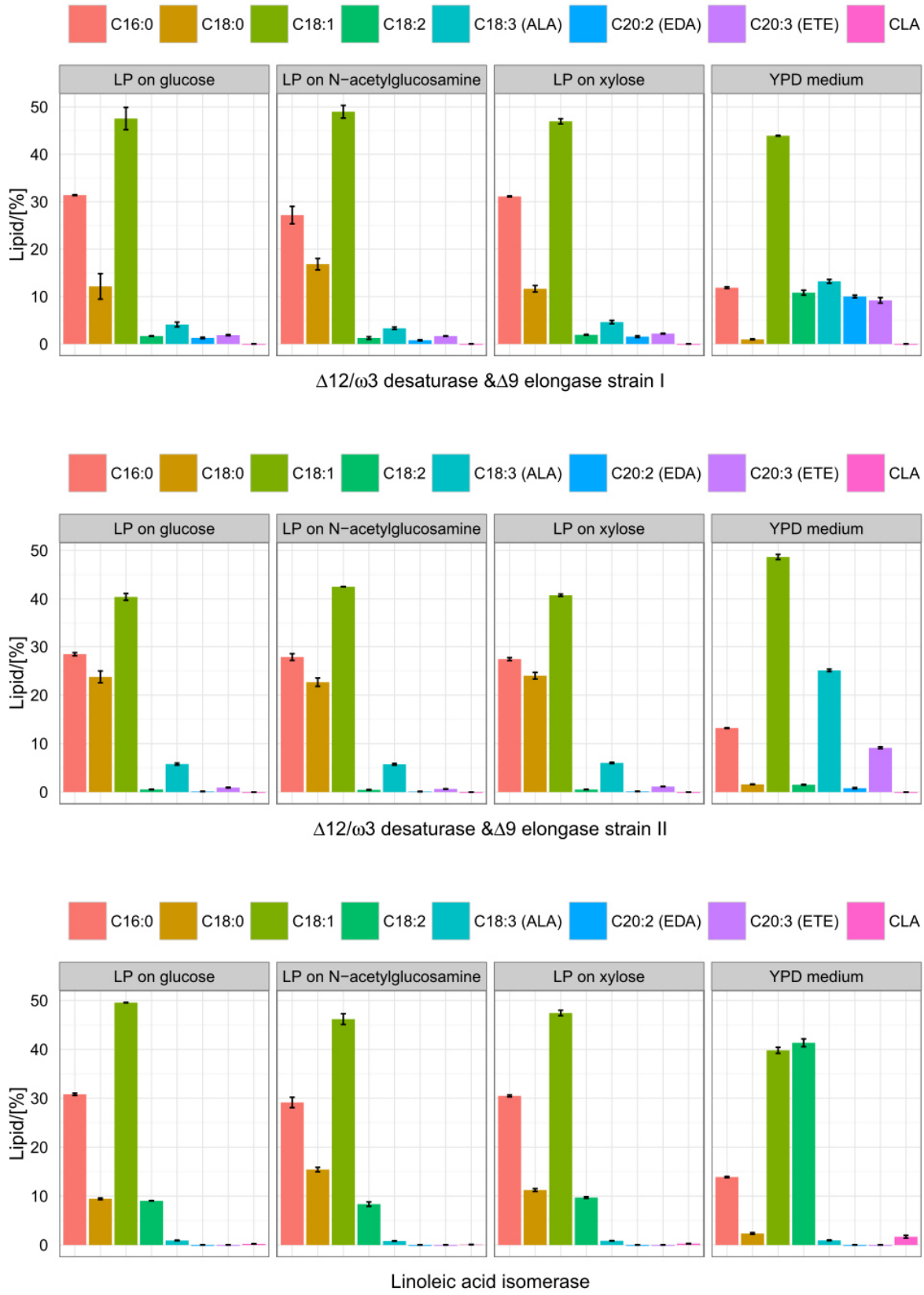
(LP = Lipid production medium with either glycose, N-acetylglucosamine or xylose as carbon source)

Figure 9. Graphical illustration of Table S1 of the fatty acid distribution after 168 hours cultivation.



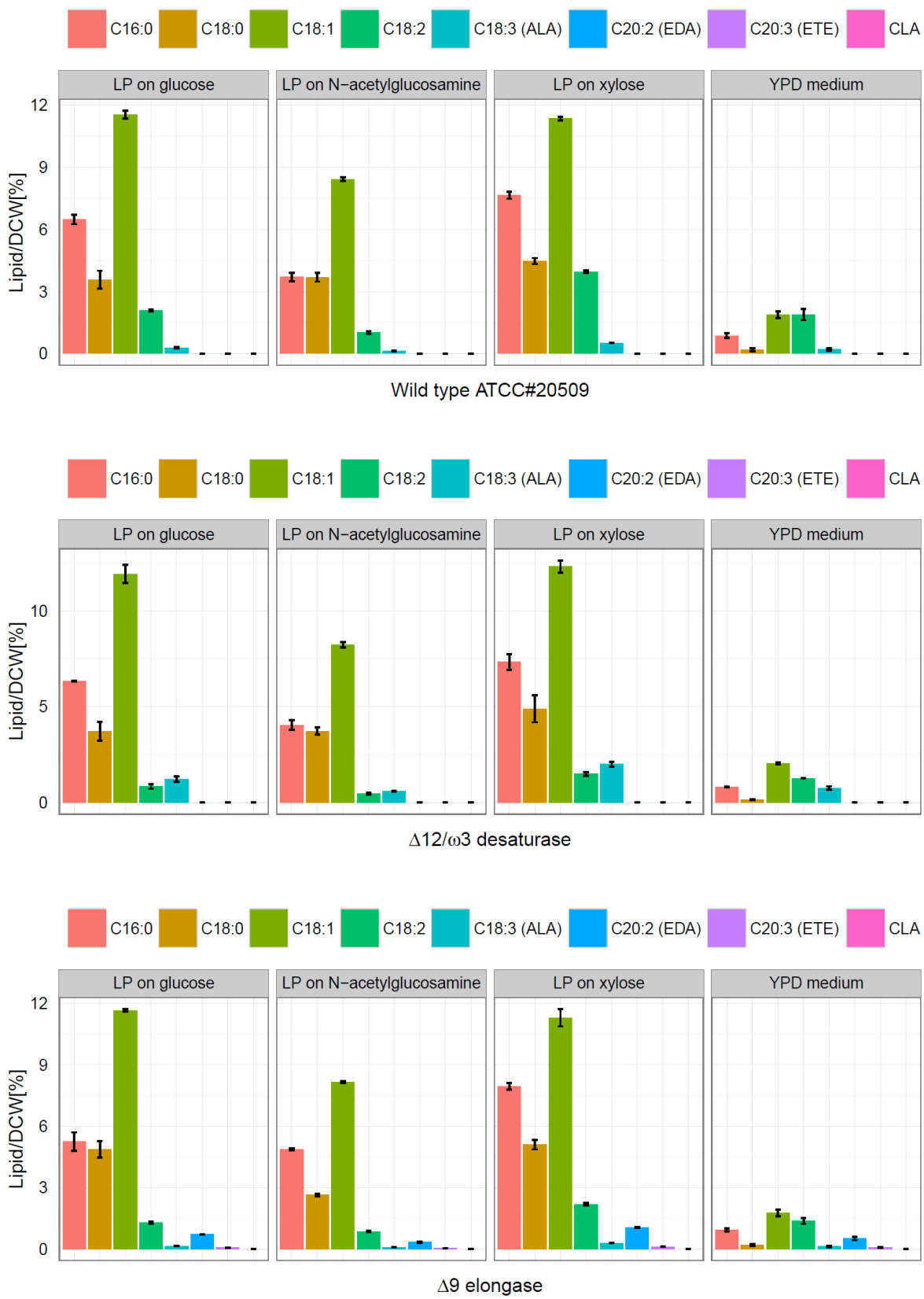
(LP = Lipid production medium with either glycose, N-acetylglucosamine, xylose as carbon source)

Figure 10. Graphical illustration of Table S1 of the fatty acid distribution after 168 hours cultivation.



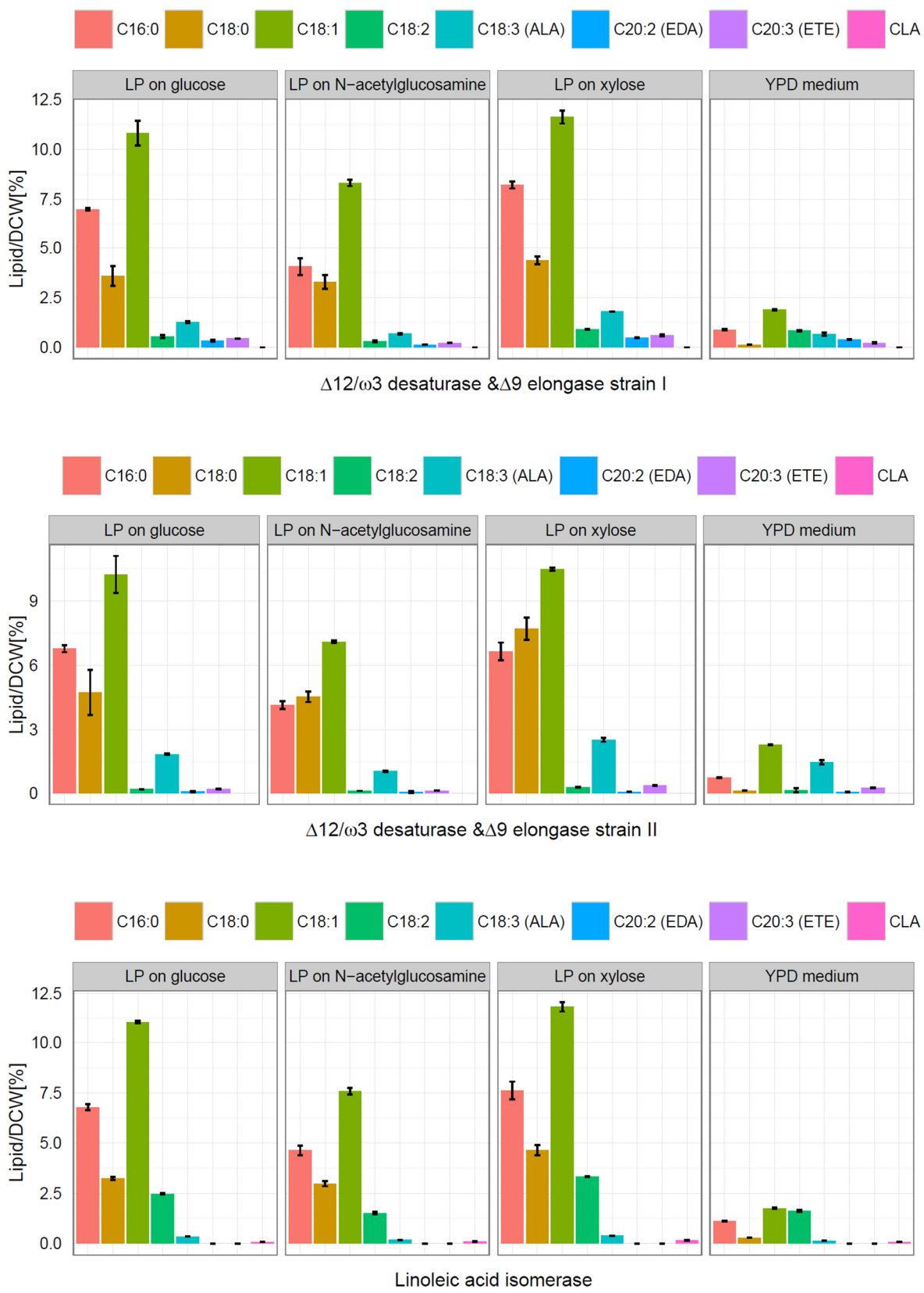
(LP = Lipid production medium with either glycose, N-acetylglucosamine, xylose as carbon source).

Figure 11. Graphical illustration of Table S2 of the fatty acid distribution after 24 hours cultivation.



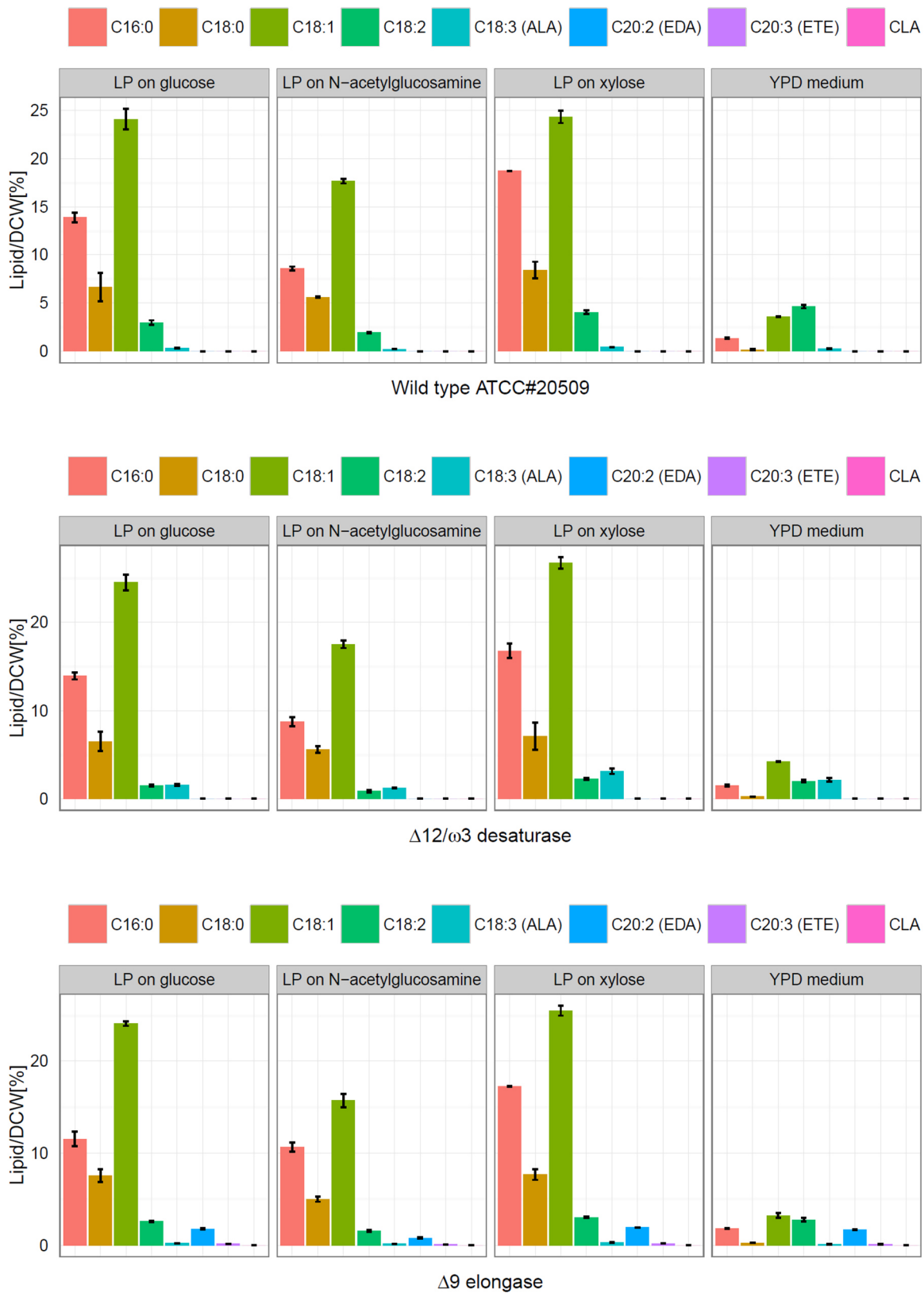
(LP = Lipid production medium with either glycose, N-acetylglucosamine, xylose as carbon source)

Figure 12. Graphical illustration of Table S2 of the fatty acid distribution after 24 hours cultivation.



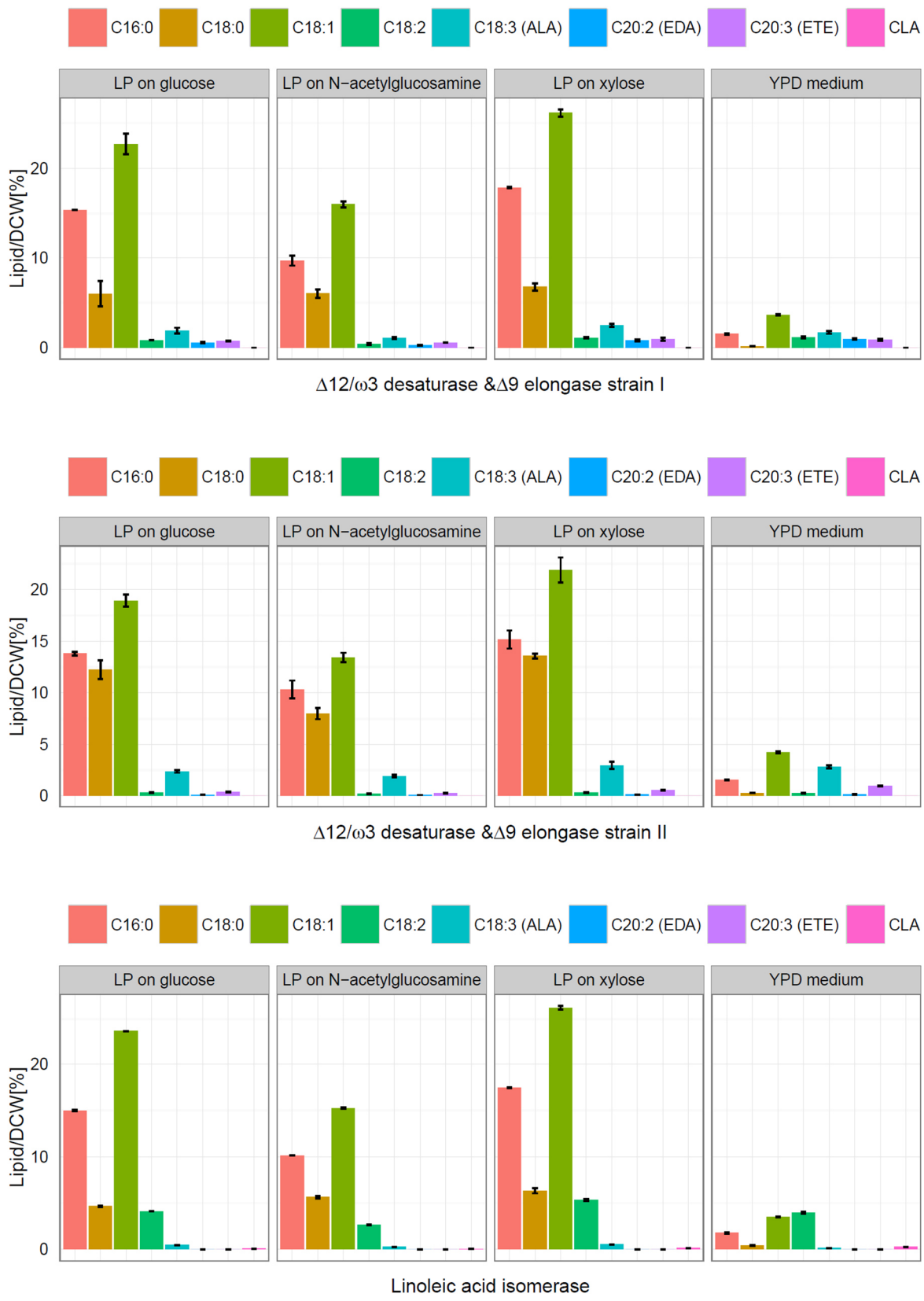
(LP = Lipid production medium with either glycose, N-acetylglucosamine, xylose as carbon source)

Figure 13. Graphical illustration of Table S2 of the fatty acid distribution after 72 hours cultivation.



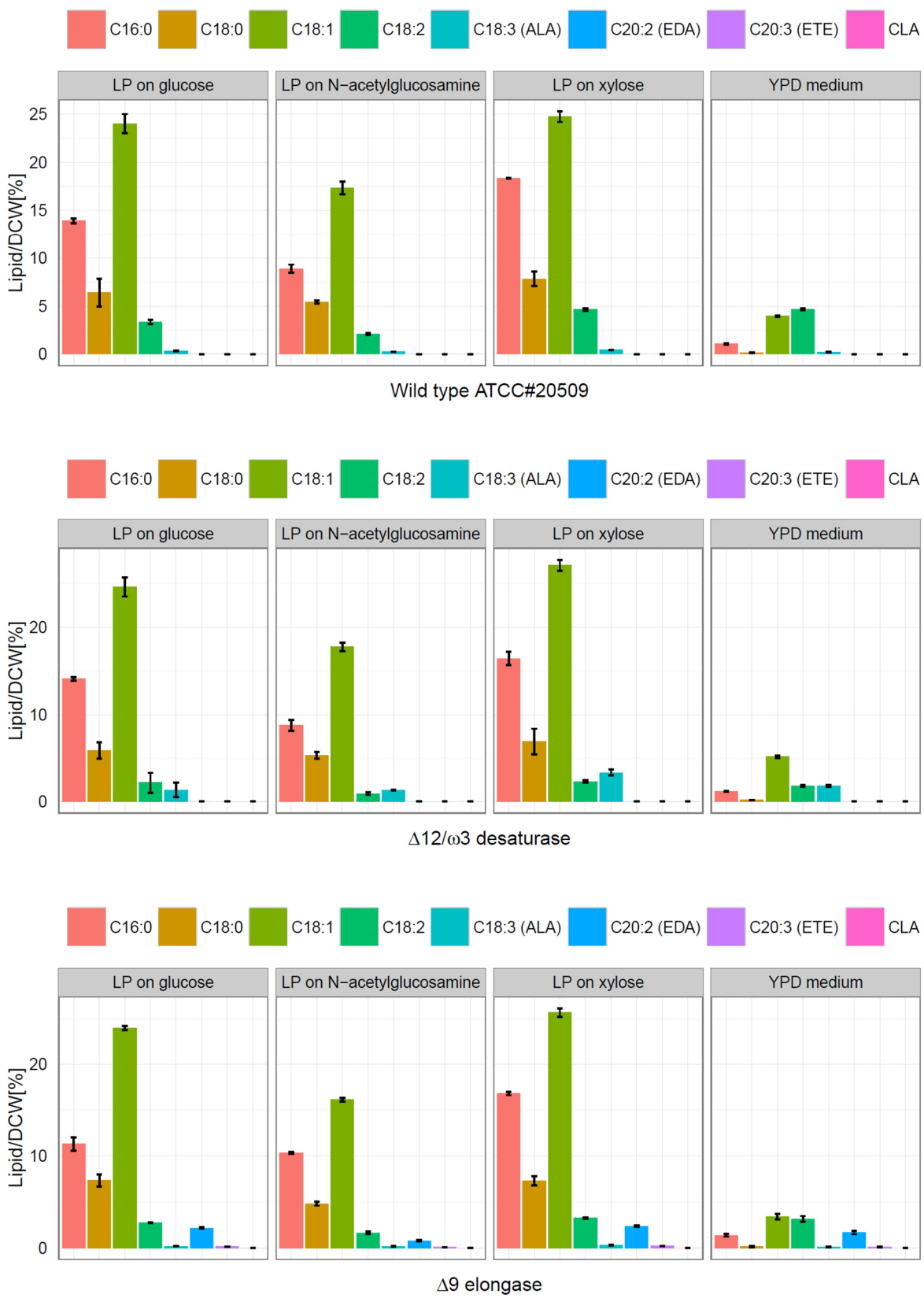
(LP = Lipid production medium with either glycose, N-acetylglucosamine, xylose as carbon source)

Figure 14. Graphical illustration of Table S2 of the fatty acid distribution after 72 hours cultivation.



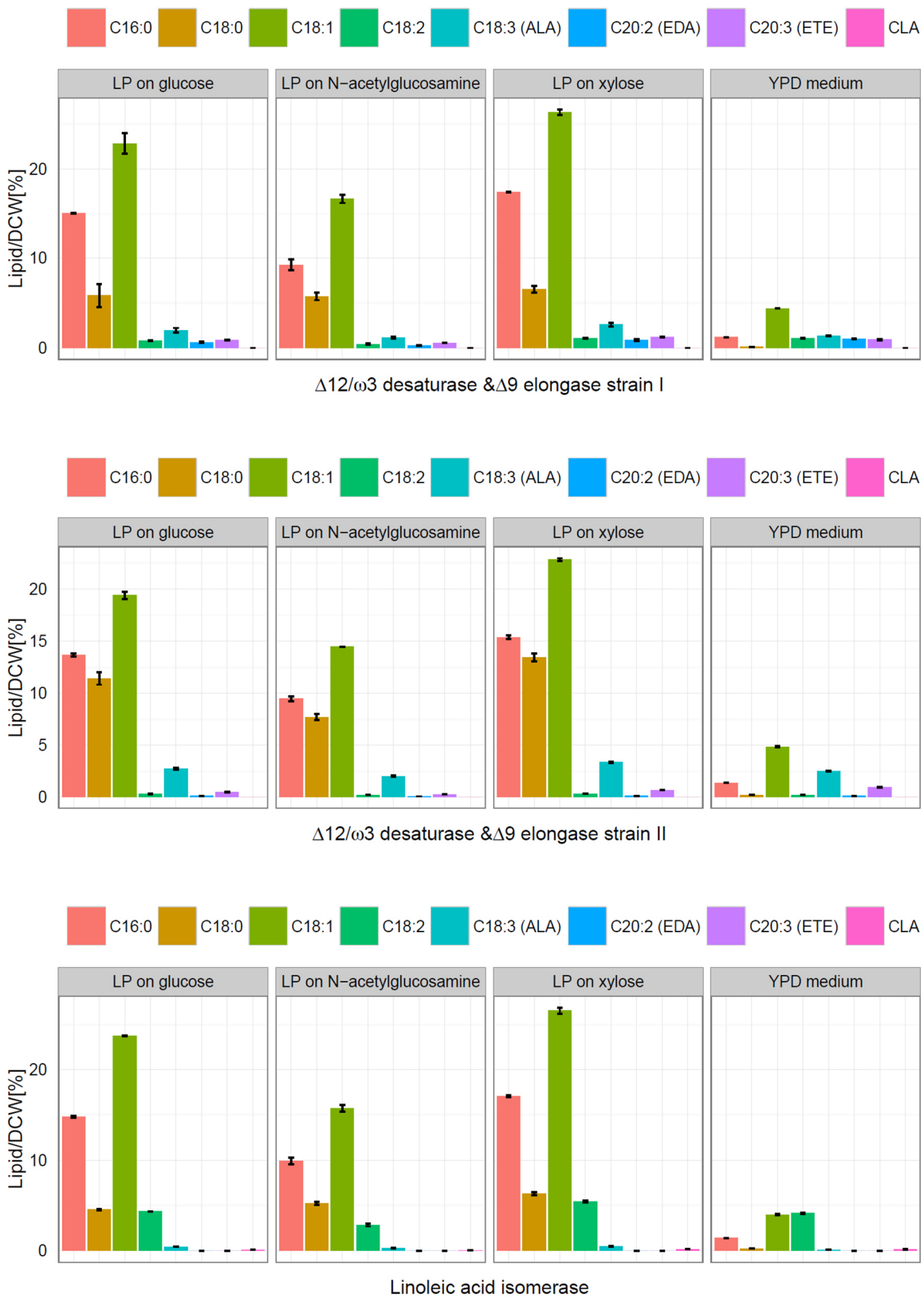
(LP = Lipid production medium with either glycose, N-acetylglucosamine, xylose as carbon source)

Figure 15. Graphical illustration of Table S2 of the fatty acid distribution after 168 hours cultivation.



(LP = Lipid production medium with either glycose, N-acetylglucosamine, xylose as carbon source)

Figure 16. Graphical illustration of Table S2 of the fatty acid distribution after 168 hours cultivation.



(LP = Lipid production medium with either glycose, N-acetylglucosamine, xylose as carbon source)

Figure 17. Growth curves of wild type *T. oleaginosus* and *T. oleaginosus expressing YFP*

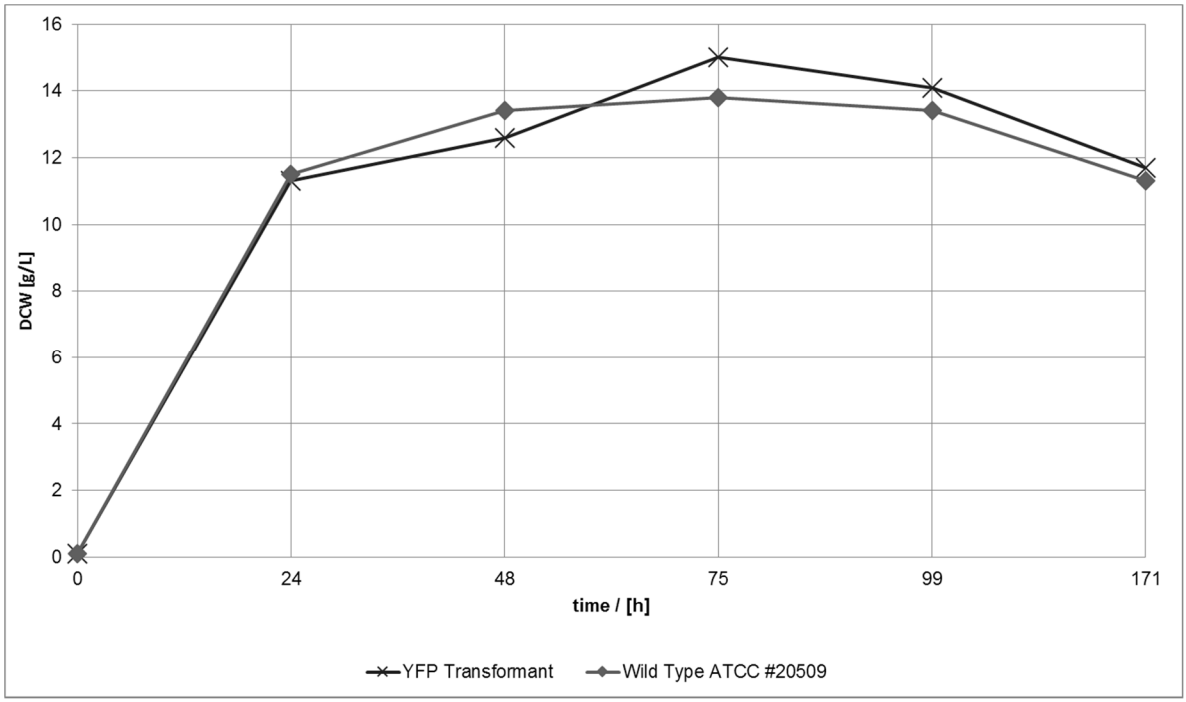


Figure 18. Yellow fluorescence protein gene expression cassette (*yfp*).

GGGGATTGGCGTCATCAAGTGCCACTGCCGCCGGCATTGGACAATGAACTTGAGACAGAGGTAAGTGCACCTCTTGCAAT
CAGGGCAGTCGATAGGCGGTGGACGAATGAGGATGCGTCGTGAGCGGGTTCGCTCCTGGCAATCTACCGCACTGGCCTG
CGTGTACTGTAGTCGCCCGCACCGCCGTGTCGGCATGGTACAAAGGCAAGAGTAGTAGTAGCACGAGCAGATGCGGGCG
TTACAGGCGTAGTAGGCGAGATCAAGCCGAACCGATGAACCCTGTCGCGTACAAGGGTCCTGATATGGGATAACCCTTCC
GGCGGCTGGGCCGGCTCGTCATGGCCCGCTCCCCCCCCCCCGGGCATGATCCCCAGATGGTCCTCTGGCCCAGTGCCTAG
GCCAACCTCCTACCTCCTCCGGCACCATTACCACCATACAGCCAGCAGCCAGCAGCCAGGCTCACCTACGGCGTCACCAG
TCGCCCTCGCTCTTGGCTACTCCTCATACCCAGCCAAACCCAGCCCATGCCAGCCACGGCCAGCCAGCCAGTCATTCTGGT
CCGTTCCATTCCCATCGATCCCTCCCCCCAGGCCCCCAACGTGTTCCGTTTCCGACCCCCACCCCGGCCTCCTCCTCCGCCC
CCCCCCCTCCGCCCTCTCCCCCCCCCATTCCAAGCCGAACCCACAACCCCATTTAACCGCGACCATCCTCGCCTCTTCCTCTC
TCTCCTTCTCTCTCTTTCTCAACCACCTCCTCTTCTCAAAACTATTCCCCTCCTCCCAAAAATCAACTTGATCAACA**ATG**GGC
AAGGTCTCGAAGGGCGAGGAGCTCTTCACCGGCGTCGTCCCCATCCTCGTCGAGCTCGACGGCGACGTCAACGGCCACA
AGTTCTCGGTCTCGGGCGAGGGCGAGGGCGACGCCACCTACGGCAAGCTCACCCTCAAGTTCATCTGCACCACCGGCAA
GCTCCCCGTCCCCTGGCCCACCCTCGTCACCACCTTCGGCTACGGCCTCCAGTGCTTCGCCCGCTACCCCGACCACATGAA
GCAGCACGACTTCTTCAAGTCGGCCATGCCCGAGGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACT
ACAAGACCCGCGCCGAGGTCAAGTTCGAGGGCGACACCCTCGTCAACCGCATCGAGCTCAAGGGCATCGACTTCAAGGA
GGACGGCAACATCCTCGGCCACAAGCTCGAGTACAACTACAACTCGCACAACGTCTACATCATGGCCGACAAGCAGAAG
AACGGCATCAAGGTCAACTTCAAGATCCGCCACAACATCGAGGACGGCTCGGTCCAGCTCGCCGACCACTACCAGCAGA
ACACCCCCATCGGCGACGGCCCCGTCCTCCTCCCCGACAACCACTACCTCTCGTACCAGTCGGCCCTCTCGAAGGACCCC
AACGAGAAGCGCGACCACATGGTCCTCCTCGAGTTCGTCACCGCCGCCGGCATCACCCTCGGCATGGACGAGCTCTACAA
G**TGA**CTTTCTAGGTTGTAGCATGGTTTCCGCGTAGCAGTAGCAGTACCAGAGGAATGCAGTTGTGACTAGATACTCACTCG
TACATGTGTGCTACCCGTGCTGACAAATGCACTGACATGGCCAGTCACTAGCTCCTCACTTCGCAATCGCAACGGCGGCGT
GTGCCCTTCAACTCCAGCGGTCATCGGCCACCGTGCCCCTCCCCTCGCCCGAATCCCCGTGCAGCGTCATCGTCCCTGCAG
GTGCGACACCCGTGTAATCATCGGGGACACCCGCGGGCCCCCTGCGAACGATGCGGAGGGCCGGCGAGTACGCGGCGCT
CCTGGCGCCGACGATGACTGAGAAAGTACCCCGGAAAGGTTATGGCGCCGTCGGGCGGAACAGCAGAAAGTATAGAATA
GCCGGGAGATATCGTGTCGGCCATGGTCGGGCAGTTACAACTCTGGACTGAAGACAGATCTGGGACCCCGCACCCCGCGT
CATCGGGTTCGGCAGGACCACGCCGACCGGAGAAATCCGGGGGACGGCTCTGAGGGCATATTCCTTGTCCAACATGTCCT
GCACCGGCTGCACCCTGCACCAAAAGGTCCAATGTCAGCGGAT

Figure 19. Δ9 elongase gene expression cassette (*igASE2*).

GGGGATTGGCGTCATCAAGTGCCACTGCCGCCGGCATTGGACAATGAACTTGAGACAGAGGTAAGTGCACCTCTTGCAAT
CAGGGCAGTCGATAGGCGGTGGACGAATGAGGATGCGTCGTGAGCGGGTTCGCTCCTGGCAATCTACCGCACTGGCCTG
CGTGTACTGTAGTCGCCCGCACCGCCGTGTCGGCATGGTACAAAGGCAAGAGTAGTAGTAGCACGAGCAGATGCGGGCG
TTACAGGCGTAGTAGGCGAGATCAAGCCGAACCGATGAACCCTGTCGCGTACAAGGGTCCTGATATGGGATAACCCTTCC
GGCGGCTGGGCCGGCTCGTCATGGCCCGCTCCCCCCCCCCCCGGGCATGATCCCCAGATGGTCCTCTGGCCCAGTGCCTAG
GCCAACCTCCTACCTCCTCCGGCACCATTACCACCATACAGCCAGCAGCCAGCAGCCAGGCTCACCTACGGCGTCACCAG
TCGCCCTCGCTCTTGGCTACTCCTCATACCCAGCCAAACCCAGCCCATGCCAGCCACGGCCAGCCAGCCAGTCATTCTGGT
CCGTTCCATTCCCATCGATCCCTCCCCCCAGGCCCCCAACGTGTTCCGTTTCCGACCCCCACCCCGGCCTCCTCCTCCGCCC
CCCCCCCTCCGCCCTCTCCCCCCCCATTCCAAGCCGAACCCACAACCCCATTTAACCGCGACCATCCTCGCCTCTTCCTCTC
TCTCCTTCTCTCTCTTTCTCAACCACCTCCTCTTCTCAAAACTATTCCCCTCCTCCCAAAAATCAACTTGATCAACA**ATG**GCC
ACCGAGGCCACCGCCTCGATCTGGGCCGCCGTCTCGGACCCCGAGATCCTCATCGGCACCTTCTCGTACCTCCTCCTCAAG
CCCATCCTCCGCTCGTCGGGCCTCGTCGACGAGAAGAAGGGCGCCTACCGCACCTCGATGATCTGGTACAACGTCATCCT
CGCCCTCTTCTCGGCCACCTCGTTCTACGTCACCGCCACCGCCCTCGGCTGGGACTACGGCTCGGGCGAGTGGCTCCGCC
GCCTCACCGGCGACACCCCCCAGCCCCTCTTCCAGTGCCCCTCGCGCGTCTGGGACTCGAAGCTCTTCGTCTGGACCGCCA
AGGCCTTCTACTACTCGAAGTACGTCGAGTACCTCGACACCGCCTGGCTCGTCCTCAAGGGCAAGAACGTCTCGTTCCTCC
AGGCCTTCCACCACTTCGGCGCCCCCTGGGACGTCTACCTCGGCATCCGCCTCCAGAACGAGGGCGTCTGGATCTTCATGT
TCTTCAACTCGTTCATCCACACCATCATGTACACCTACTACGGCCTCACCGCCGCCGGCTACAAGATCAAGGCCAAGCCCC
TCATCACCGCCATGCAGATCTCGCAGTTCATGGGCGGCTTCATCCTCGTCTGGGACTACATCAACATCCCCTGCTTCCGCTC
GGACAACGGCAAGGTCTTCTCGTGGGTCTTCAACTACGCCTACGTCGGCTTCGTCTTCCTCCTCTTCTGCCACTTCTTCTACA
AGGACAACCTCGCCTCGAAGAAGCCCGCCAAGGGCGGCAAGGCCCTC**TGA**CTTTCTAGGTTGTAGCATGGTTTCCGCGTA
GCAGTAGCAGTACCAGAGGAATGCAGTTGTGACTAGATACTCACTCGTACATGTGTGCTACCCGTGCTGACAAATGCACTG
ACATGGCCAGTCACTAGCTCCTCACTTCGCAATCGCAACGGCGGCGTGTGCCCTTCAACTCCAGCGGTCATCGGCCACCGT
GCCCCTCCCCTCGCCCGAATCCCCGTGCAGCGTCATCGTCCCTGCAGGTGCGACACCCGTGTAATCATCGGGGACACCCG
CGGGCCCCCTGCGAACGATGCGGAGGGCCGGCGAGTACGCGGCGCTCCTGGCGCCGACGATGACTGAGAAAGTACCCC
GGAAAGGTTATGGCGCCGTCGGGCGGAACAGCAGAAAGTATAGAATAGCCGGGAGATATCGTGTCGGCCATGGTCGGGC
AGTTACAACTCTGGACTGAAGACAGATCTGGGACCCCGCACCCCGCGTCATCGGGTTCGGCAGGACCACGCCGACCGGA
GAAATCCGGGGGACGGCTCTGAGGGCATATTCCTTGTCCAACATGTCCTGCACCGGCTGCACCCTGCACCAAAAGGTCCA
ATGTCAGCGGAT

Figure 20. Δ12 /ω3 desaturase gene expression cassette (*fm1*).

GGGGATTGGCGTCATCAAGTGCCACTGCCGCCGGCATTGGACAATGAACTTGAGACAGAGGTAAGTGCACCTCTTGCAAT
CAGGGCAGTCGATAGGCGGTGGACGAATGAGGATGCGTCGTGAGCGGGTTCGCTCCTGGCAATCTACCGCACTGGCCTG
CGTGTACTGTAGTCGCCCGCACCGCCGTGTCGGCATGGTACAAAGGCAAGAGTAGTAGTAGCACGAGCAGATGCGGGCG
TTACAGGCGTAGTAGGCGAGATCAAGCCGAACCGATGAACCCTGTCGCGTACAAGGGTCCTGATATGGGATAACCCTTCC
GGCGGCTGGGCCGGCTCGTCATGGCCCGCTCCCCCCCCCCCCGGGCATGATCCCCAGATGGTCCTCTGGCCCAGTGCCTAG
GCCAACCTCCTACCTCCTCCGGCACCATTACCACCATACAGCCAGCAGCCAGCAGCCAGGCTCACCTACGGCGTCACCAG
TCGCCCTCGCTCTTGGCTACTCCTCATACCCAGCCAAACCCAGCCCATGCCAGCCACGGCCAGCCAGCCAGTCATTCTGGT
CCGTTCCATTCCCATCGATCCCTCCCCCCAGGCCCCCAACGTGTTCCGTTTCCGACCCCCACCCCGGCCTCCTCCTCCGCCC
CCCCCCCTCCGCCCTCTCCCCCCCCCATTCCAAGCCGAACCCACAACCCCATTTAACCGCGACCATCCTCGCCTCTTCCTCTC
TCTCCTTCTCTCTCTTTCTCAACCACCTCCTCTTCTCAAAACTATTCCCCTCCTCCCAAAAATCAACTTGATCAACA**ATG**GCC
ACCCGCCAGCGCACCGCCACCACCGTCGTCGTCGAGGACCTCCCCAAGGTCACCCTCGAGGCCAAGTCGGAGCCCGTCTT
CCCCGACATCAAGACCATCAAGGACGCCATCCCCGCCCACTGCTTCCAGCCCTCGCTCGTCACCTCGTTCTACTACGTCTTC
CGCGACTTCGCCATGGTCTCGGCCCTCGTCTGGGCCGCCCTCACCTACATCCCCTCGATCCCCGACCAGACCCTCCGCGTC
GCCGCCTGGATGGTCTACGGCTTCGTCCAGGGCCTCTTCTGCACCGGCGTCTGGATCCTCGGCCACGAGTGCGGCCACGG
CGCCTTCTCGCTCCACGGCAAGGTCAACAACGTCACCGGCTGGTTCCTCCACTCGTTCCTCCTCGTCCCCTACTTCTCGTGG
AAGTACTCGCACCACCGCCACCACCGCTTCACCGGCCACATGGACCTCGACATGGCCTTCGTCCCCAAGACCGAGCCCAA
GCCCTCGAAGTCGCTCATGATCGCCGGCATCGACGTCGCCGAGCTCGTCGAGGACACCCCCGCCGCCCAGATGGTCAAGC
TCATCTTCCACCAGCTCTTCGGCTGGCAGGCCTACCTCTTCTTCAACGCCTCGTCGGGCAAGGGCTCGAAGCAGTGGGAGC

CCAAGACCGGCCTCTCGAAGTGGTTCCGCGTCTCGCACTTCGAGCCCACCTCGGCCGTCTTCCGCCCCAACGAGGCCATCT
TCATCCTCATCTCGGACATCGGCCTCGCCCTCATGGGCACCGCCCTCTACTTCGCCTCGAAGCAGGTCGGCGTCTCGACCA
TCCTCTTCCTCTACCTCGTCCCCTACCTCTGGGTCCACCACTGGCTCGTCGCCATCACCTACCTCCACCACCACCACACCGA
GCTCCCCCACTACACCGCCGAGGGCTGGACCTACGTCAAGGGCGCCCTCGCCACCGTCGACCGCGAGTTCGGCTTCATCG
GCAAGCACCTCTTCCACGGCATCATCGAGAAGCACGTCGTCCACCACCTCTTCCCCAAGATCCCCTTCTACAAGGCCGACG
AGGCCACCGAGGCCATCAAGCCCGTCATCGGCGACCACTACTGCCACGACGACCGCTCGTTCCTCGGCCAGCTCTGGACC
ATCTTCGGCACCCTCAAGTACGTCGAGCACGACCCCGCCCGCCCCGGCGCCATGCGCTGGAACAAGGAC**TGA**CTTTCTAG
GTTGTAGCATGGTTTCCGCGTAGCAGTAGCAGTACCAGAGGAATGCAGTTGTGACTAGATACTCACTCGTACATGTGTGCT
ACCCGTGCTGACAAATGCACTGACATGGCCAGTCACTAGCTCCTCACTTCGCAATCGCAACGGCGGCGTGTGCCCTTCAAC
TCCAGCGGTCATCGGCCACCGTGCCCCTCCCCTCGCCCGAATCCCCGTGCAGCGTCATCGTCCCTGCAGGTGCGACACCC
GTGTAATCATCGGGGACACCCGCGGGCCCCCTGCGAACGATGCGGAGGGCCGGCGAGTACGCGGCGCTCCTGGCGCCG
ACGATGACTGAGAAAGTACCCCGGAAAGGTTATGGCGCCGTCGGGCGGAACAGCAGAAAGTATAGAATAGCCGGGAGAT
ATCGTGTCGGCCATGGTCGGGCAGTTACAACTCTGGACTGAAGACAGATCTGGGACCCCGCACCCCGCGTCATCGGGTTC
GGCAGGACCACGCCGACCGGAGAAATCCGGGGGACGGCTCTGAGGGCATATTCCTTGTCCAACATGTCCTGCACCGGCT
GCACCCTGCACCAAAAGGTCCAATGTCAGCGGAT

Figure 21. Linoleic acid isomerase expression cassette (*pAI*).

GGGGATTGGCGTCATCAAGTGCCACTGCCGCCGGCATTGGACAATGAACTTGAGACAGAGGTAAGTGCACCTCTTGCAAT
CAGGGCAGTCGATAGGCGGTGGACGAATGAGGATGCGTCGTGAGCGGGTTCGCTCCTGGCAATCTACCGCACTGGCCTG
CGTGTACTGTAGTCGCCCGCACCGCCGTGTCGGCATGGTACAAAGGCAAGAGTAGTAGTAGCACGAGCAGATGCGGGCG
TTACAGGCGTAGTAGGCGAGATCAAGCCGAACCGATGAACCCTGTCGCGTACAAGGGTCCTGATATGGGATAACCCTTCC
GGCGGCTGGGCCGGCTCGTCATGGCCCGCTCCCCCCCCCCCGGGCATGATCCCCAGATGGTCCTCTGGCCCAGTGCCTAG
GCCAACCTCCTACCTCCTCCGGCACCATTACCACCATACAGCCAGCAGCCAGCAGCCAGGCTCACCTACGGCGTCACCAG
TCGCCCTCGCTCTTGGCTACTCCTCATACCCAGCCAAACCCAGCCCATGCCAGCCACGGCCAGCCAGCCAGTCATTCTGGT
CCGTTCCATTCCCATCGATCCCTCCCCCCAGGCCCCCAACGTGTTCCGTTTCCGACCCCCACCCCGGCCTCCTCCTCCGCCC
CCCCCCCTCCGCCCTCTCCCCCCCCCATTCCAAGCCGAACCCACAACCCCATTTAACCGCGACCATCCTCGCCTCTTCCTCTC
TCTCCTTCTCTCTCTTTCTCAACCACCTCCTCTTCTCAAAACTATTCCCCTCCTCCCAAAAATCAACTTGATCAACA**ATG**TCGA
TCTCGAAGGACTCGCGCATCGCCATCATCGGCGCCGGCCCCGCCGGCCTCGCCGCCGGCATGTACCTCGAGCAGGCCGG
CTTCCACGACTACACCATCCTCGAGCGCACCGACCACGTCGGCGGCAAGTGCCACTCGCCCAACTACCACGGCCGCCGCT
ACGAGATGGGCGCCATCATGGGCGTCCCCTCGTACGACACCATCCAGGAGATCATGGACCGCACCGGCGACAAGGTCGA
CGGCCCCAAGCTCCGCCGCGAGTTCCTCCACGAGGACGGCGAGATCTACGTCCCCGAGAAGGACCCCGTCCGCGGCCCC
CAGGTCATGGCCGCCGTCCAGAAGCTCGGCCAGCTCCTCGCCACCAAGTACCAGGGCTACGACGCCAACGGCCACTACA
ACAAGGTCCACGAGGACCTCATGCTCCCCTTCGACGAGTTCCTCGCCCTCAACGGCTGCGAGGCCGCCCGCGACCTCTGG
ATCAACCCCTTCACCGCCTTCGGCTACGGCCACTTCGACAACGTCCCCGCCGCCTACGTCCTCAAGTACCTCGACTTCGTC
ACCATGATGTCGTTCGCCAAGGGCGACCTCTGGACCTGGGCCGACGGCACCCAGGCCATGTTCGAGCACCTCAACGCCAC
CCTCGAGCACCCCGCCGAGCGCAACGTCGACATCACCCGCATCACCCGCGAGGACGGCAAGGTCCACATCCACACCACC
GACTGGGACCGCGAGTCGGACGTCCTCGTCCTCACCGTCCCCCTCGAGAAGTTCCTCGACTACTCGGACGCCGACGACGA
CGAGCGCGAGTACTTCTCGAAGATCATCCACCAGCAGTACATGGTCGACGCCTGCCTCGTCAAGGAGTACCCCACCATCT
CGGGCTACGTCCCCGACAACATGCGCCCCGAGCGCCTCGGCCACGTCATGGTCTACTACCACCGCTGGGCCGACGACCCC
CACCAGATCATCACCACCTACCTCCTCCGCAACCACCCCGACTACGCCGACAAGACCCAGGAGGAGTGCCGCCAGATGGT
CCTCGACGACATGGAGACCTTCGGCCACCCCGTCGAGAAGATCATCGAGGAGCAGACCTGGTACTACTTCCCCCACGTCT
CGTCGGAGGACTACAAGGCCGGCTGGTACGAGAAGGTCGAGGGCATGCAGGGCCGCCGCAACACCTTCTACGCCGGCG
AGATCATGTCGTTCGGCAACTTCGACGAGGTCTGCCACTACTCGAAGGACCTCGTCACCCGCTTCTTCGTC**TGA**CTTTCTAG
GTTGTAGCATGGTTTCCGCGTAGCAGTAGCAGTACCAGAGGAATGCAGTTGTGACTAGATACTCACTCGTACATGTGTGCT
ACCCGTGCTGACAAATGCACTGACATGGCCAGTCACTAGCTCCTCACTTCGCAATCGCAACGGCGGCGTGTGCCCTTCAAC
TCCAGCGGTCATCGGCCACCGTGCCCCTCCCCTCGCCCGAATCCCCGTGCAGCGTCATCGTCCCTGCAGGTGCGACACCC
GTGTAATCATCGGGGACACCCGCGGGCCCCCTGCGAACGATGCGGAGGGCCGGCGAGTACGCGGCGCTCCTGGCGCCG
ACGATGACTGAGAAAGTACCCCGGAAAGGTTATGGCGCCGTCGGGCGGAACAGCAGAAAGTATAGAATAGCCGGGAGAT
ATCGTGTCGGCCATGGTCGGGCAGTTACAACTCTGGACTGAAGACAGATCTGGGACCCCGCACCCCGCGTCATCGGGTTC
GGCAGGACCACGCCGACCGGAGAAATCCGGGGGACGGCTCTGAGGGCATATTCCTTGTCCAACATGTCCTGCACCGGCT
GCACCCTGCACCAAAAGGTCCAATGTCAGCGGAT

Figure 22. Mass spectrum of cembrene A recorded on a Trace GC Ultra with DSQII (Thermo Scientific), *m/z* was analysed from [50-650].

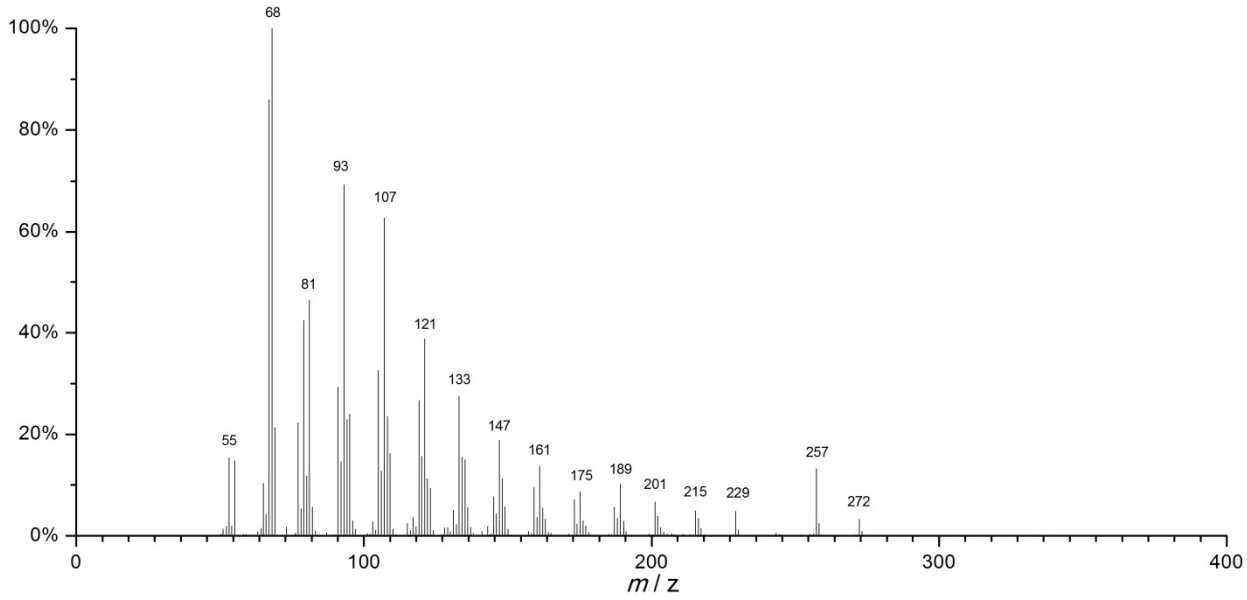


Figure 23. Mass spectrum of cyclooctat-7-en-3-ol recorded on a DFS high resolution GC/MS (Thermo Scientific), *m/z* was analyzed from [35-500].

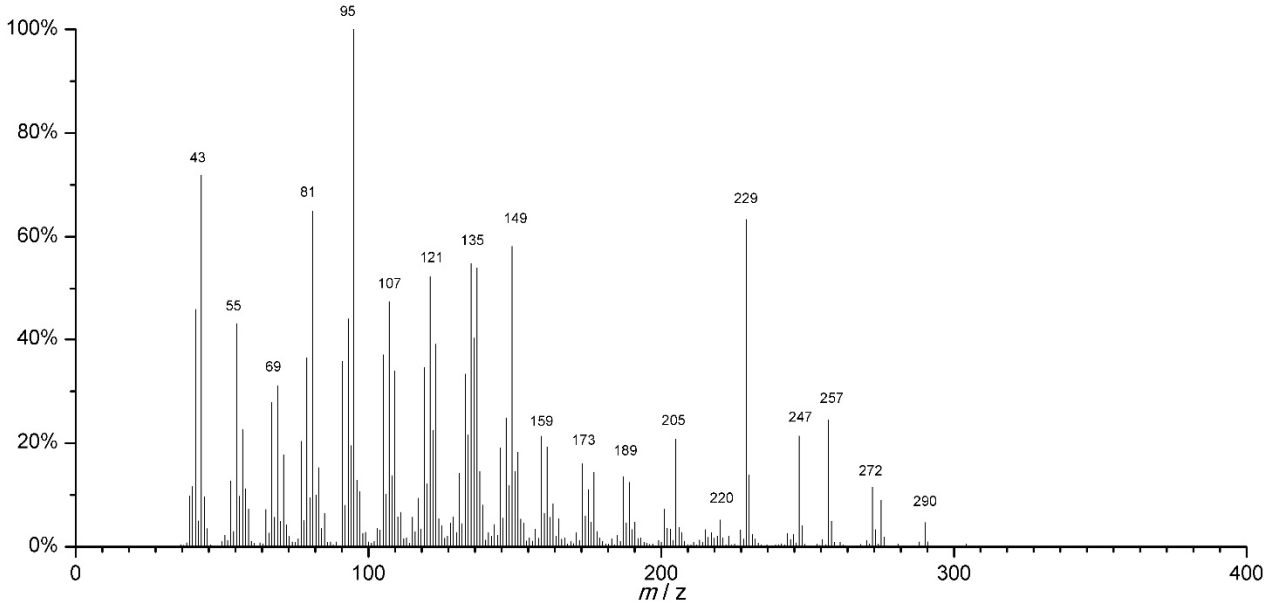


Figure 24. Mass spectrum of the diterpene product F107Y_A, recorded on a DFS high resolution GC/MS (Thermo Scientific), *m/z* was analyzed from [35-500].

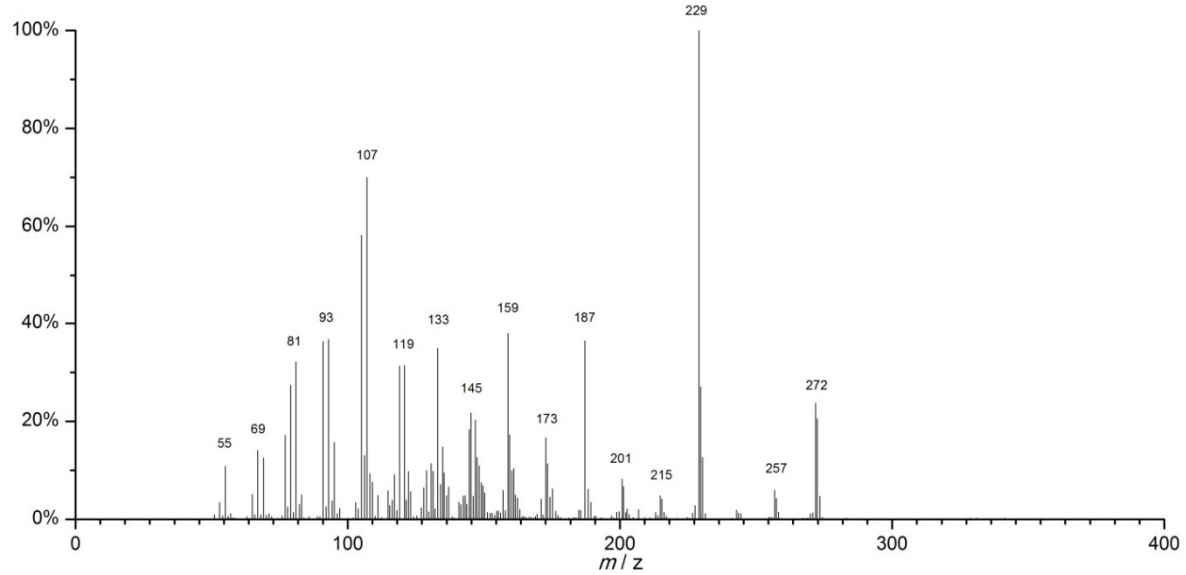


Figure 25. Mass spectrometery of Cyclooctat-1,7-diene (F107Y_B), recorded on a DFS high resolution GC/MS (Thermo Scientific), *m/z* was analyzed from [35-500].

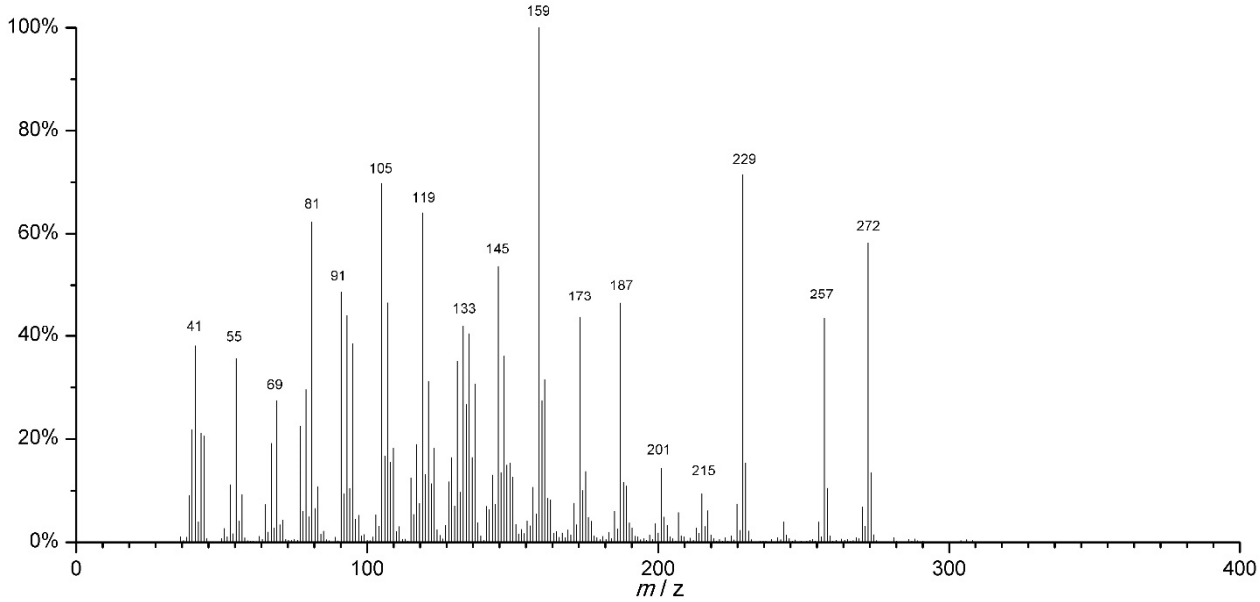


Figure 26. Mass spectrum of (1R,3E,7E,11S,12S)-3,7,18-dolabellatriene recorded on a Trace GC Ultra with DSQII (Thermo Scientific), *m/z* was analyzed from [50-650].

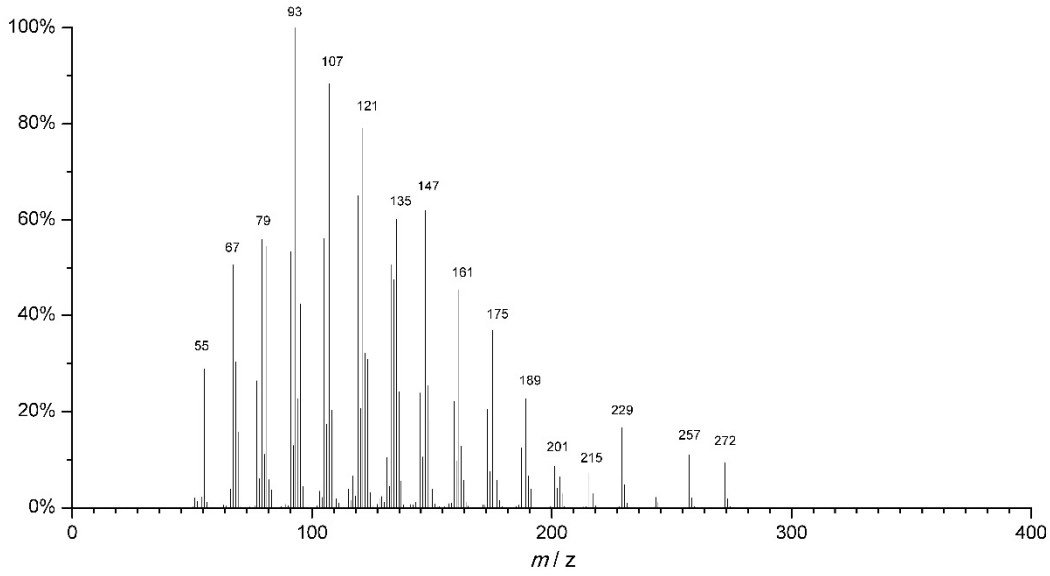


Figure 27. 1H NMR (500 MHz, CDCl3) of cyclooctat-9-en-7-ol

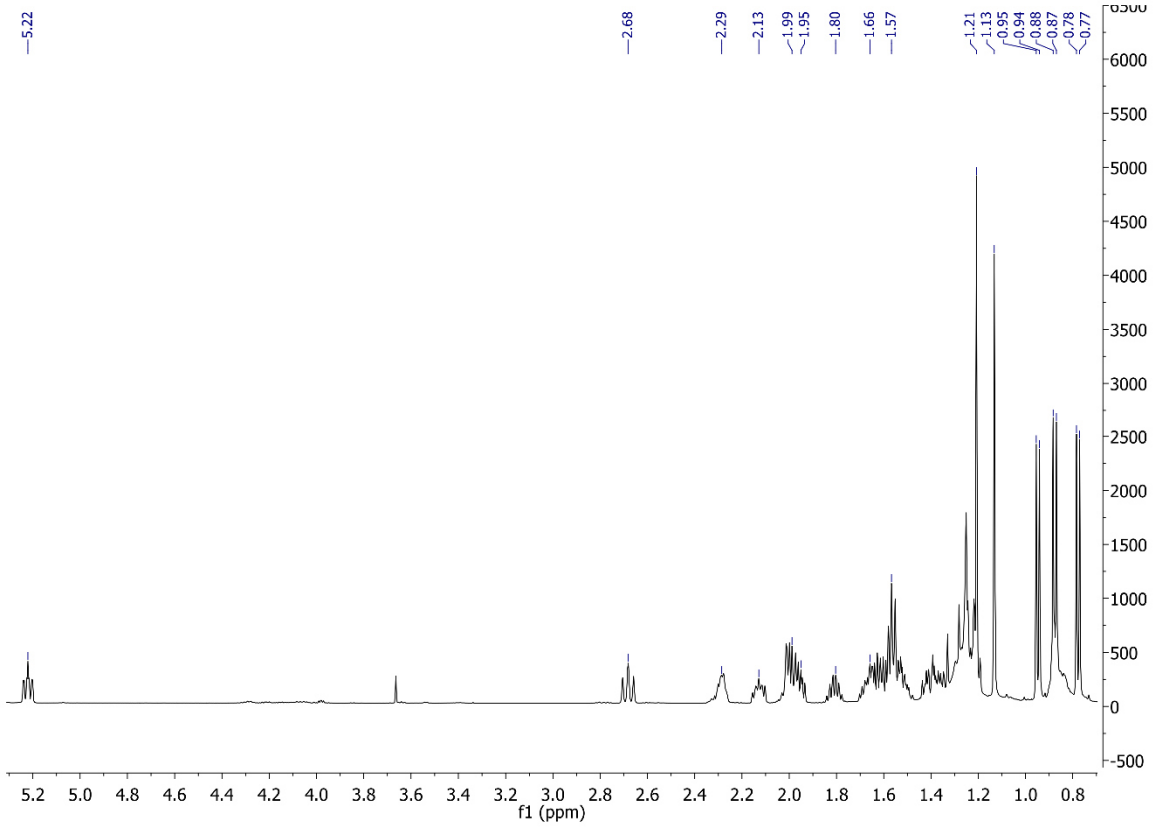


Figure 28. ^{13}C NMR (125 MHz, $CDCl_3$) of cyclooctat-9-en-7-ol

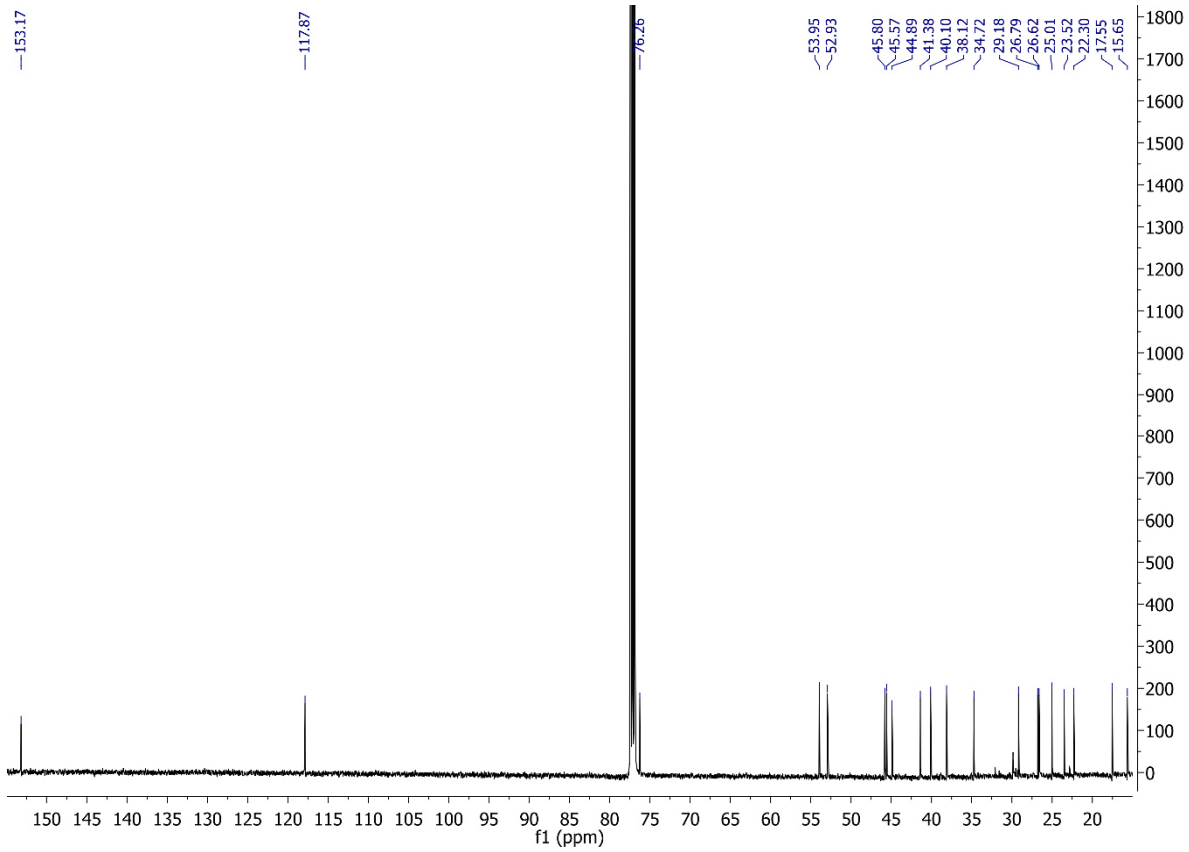


Figure 29. ^{1}H NMR (500 MHz, CDCl3) of R-cembrene A

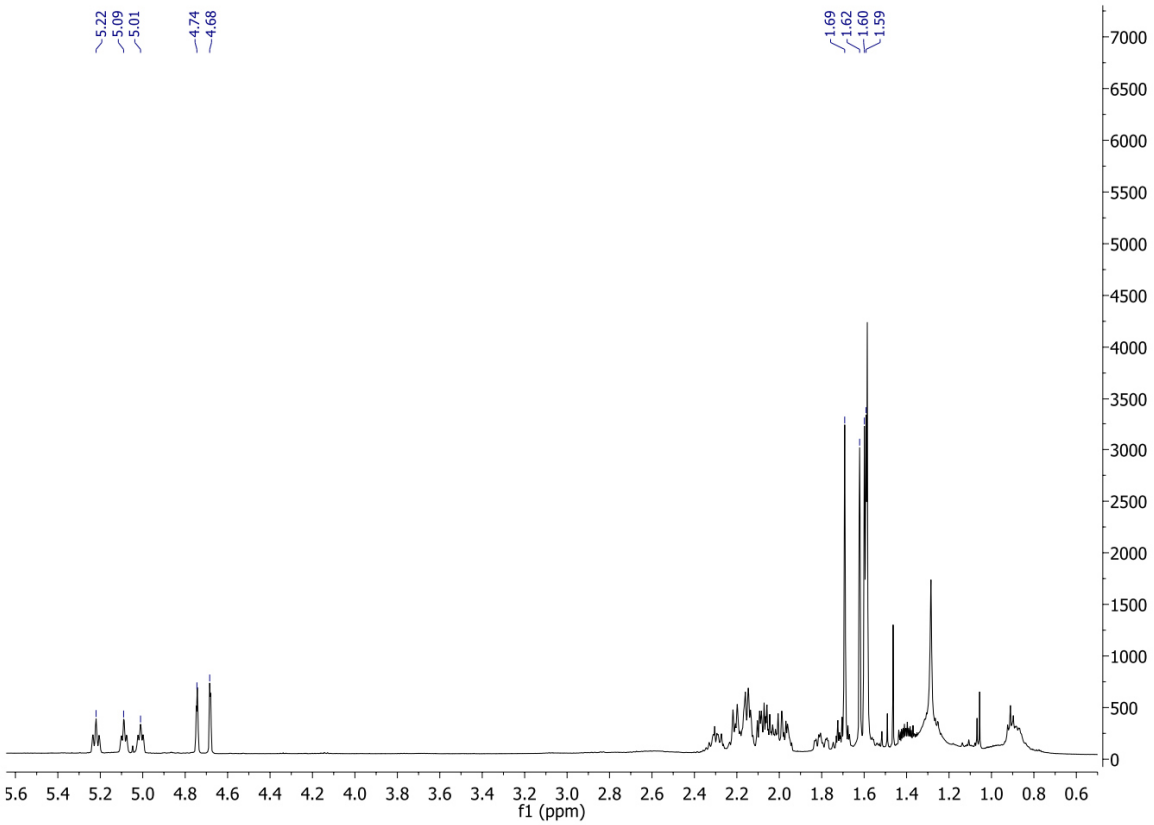


Figure 30. ^{13}C NMR (125 MHz, $CDCl_3$) of R-cembrene A

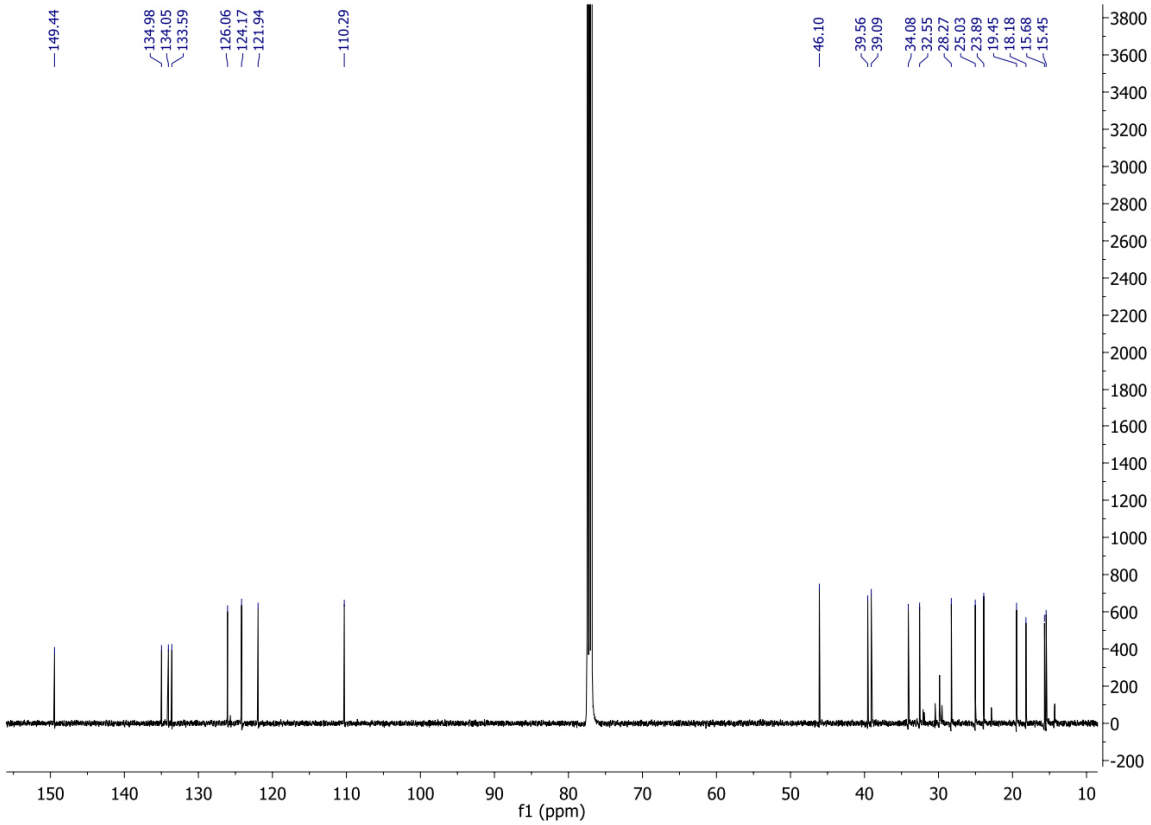


Figure 31 ^{1}H NMR (500 MHz, CDCl3) of cyclooctat-7-en-3-ol

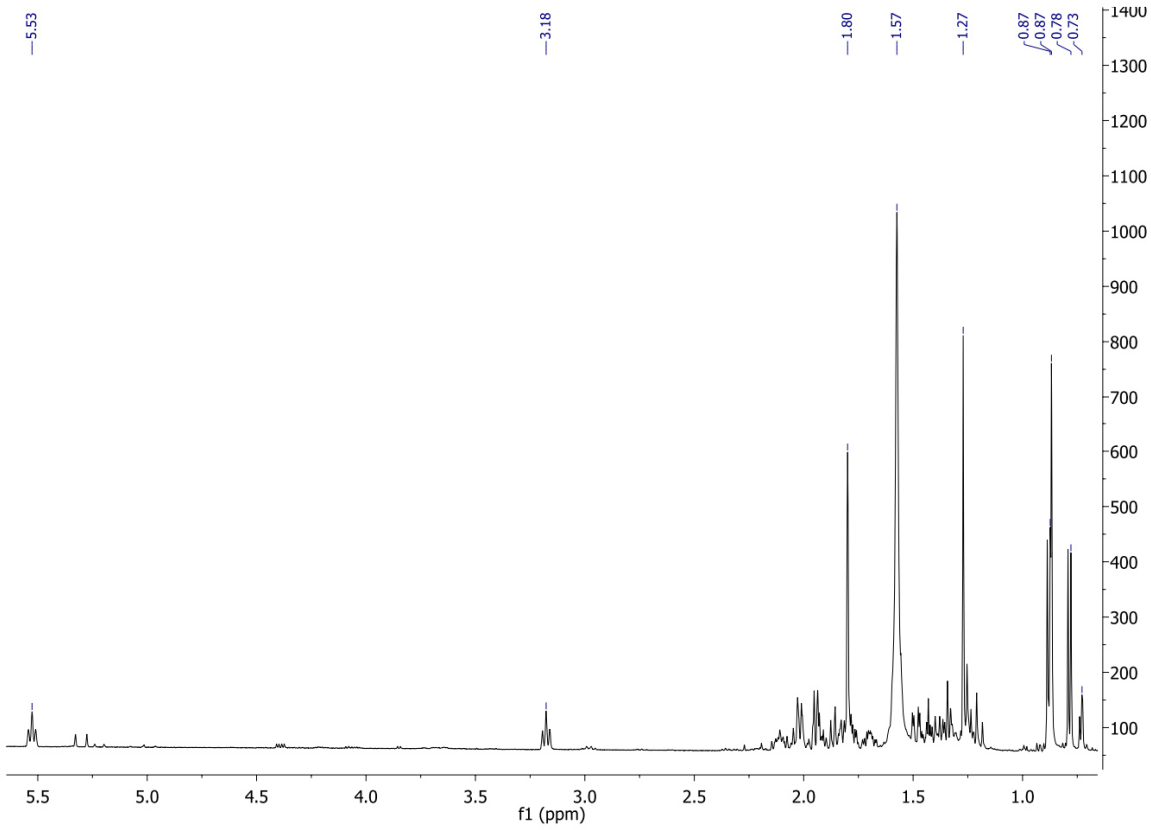


Figure 32. ^{13}C NMR (125 MHz, $CDCl_3$) of cyclooctat-7-en-3-ol

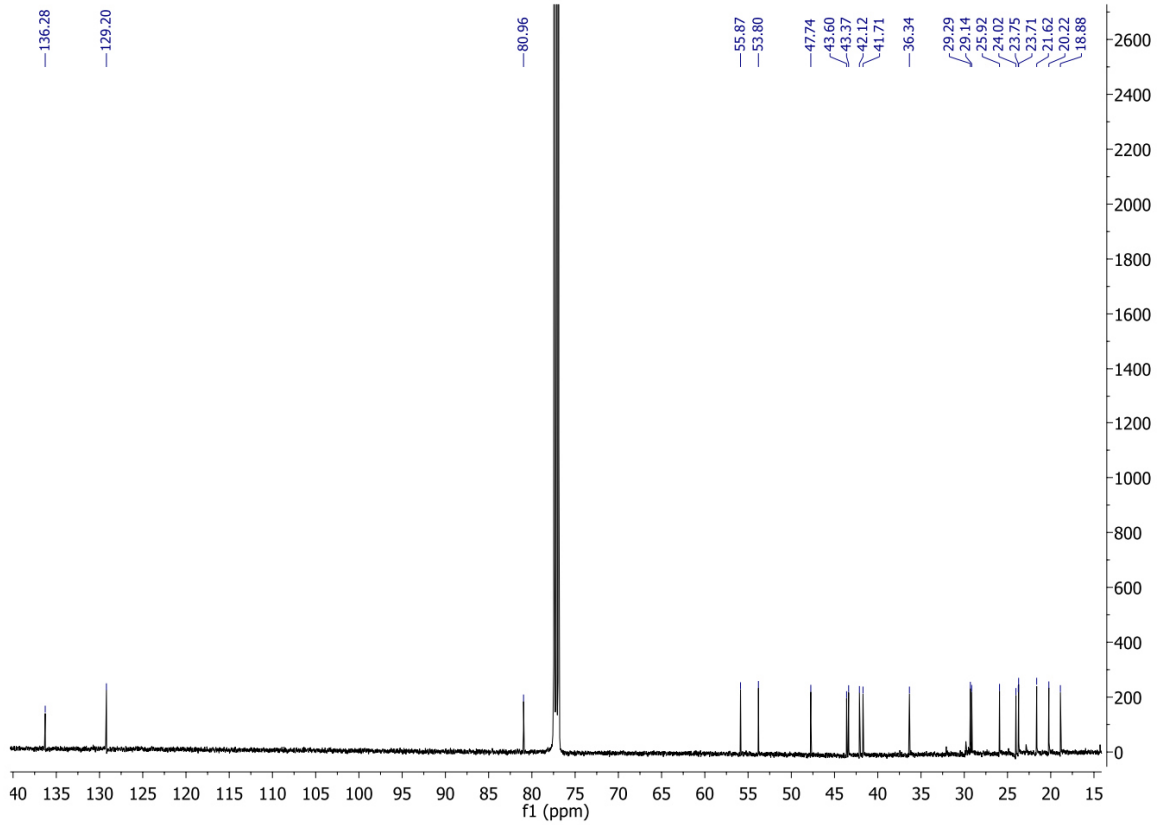


Figure 33. COSY NMR (500 MHz, CDCl3) of cyclooctat-7-en-3-ol

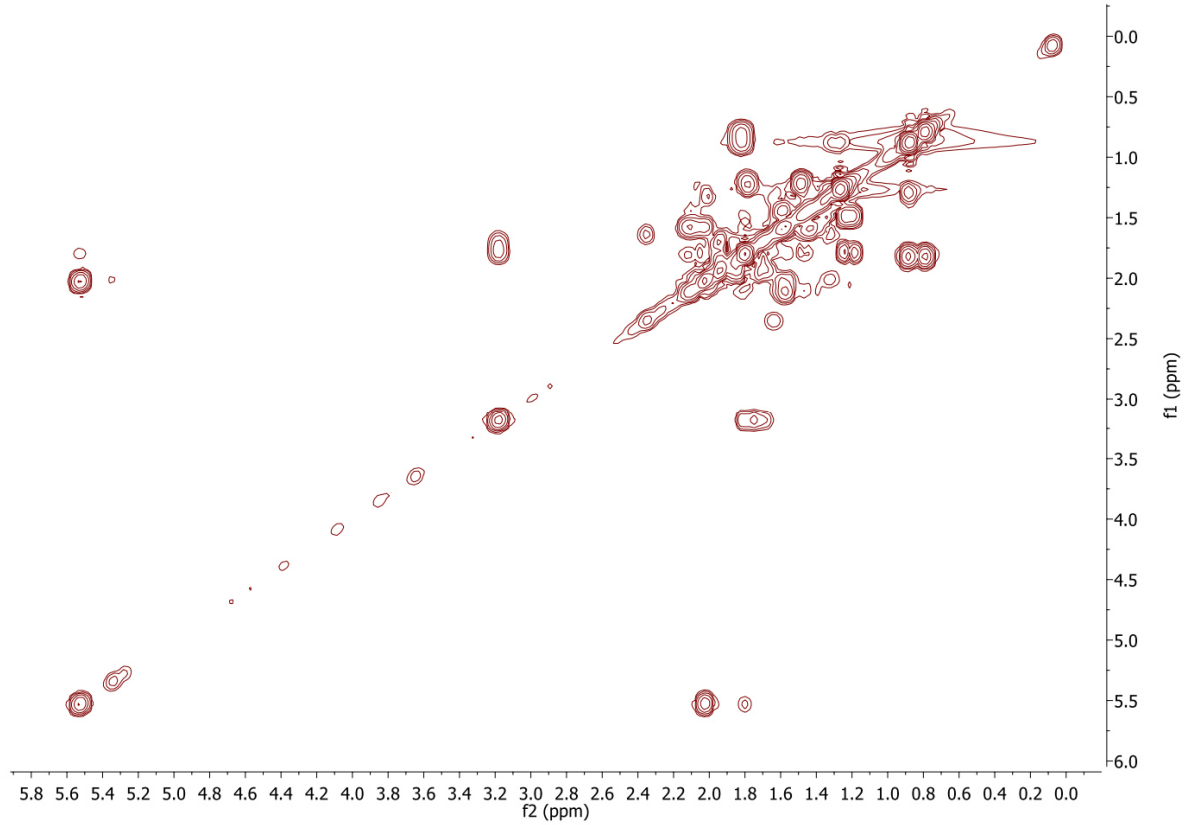


Figure 34. HSQC NMR (500 MHz, CDCl3) of Cyclooctat-7-en-3-ol

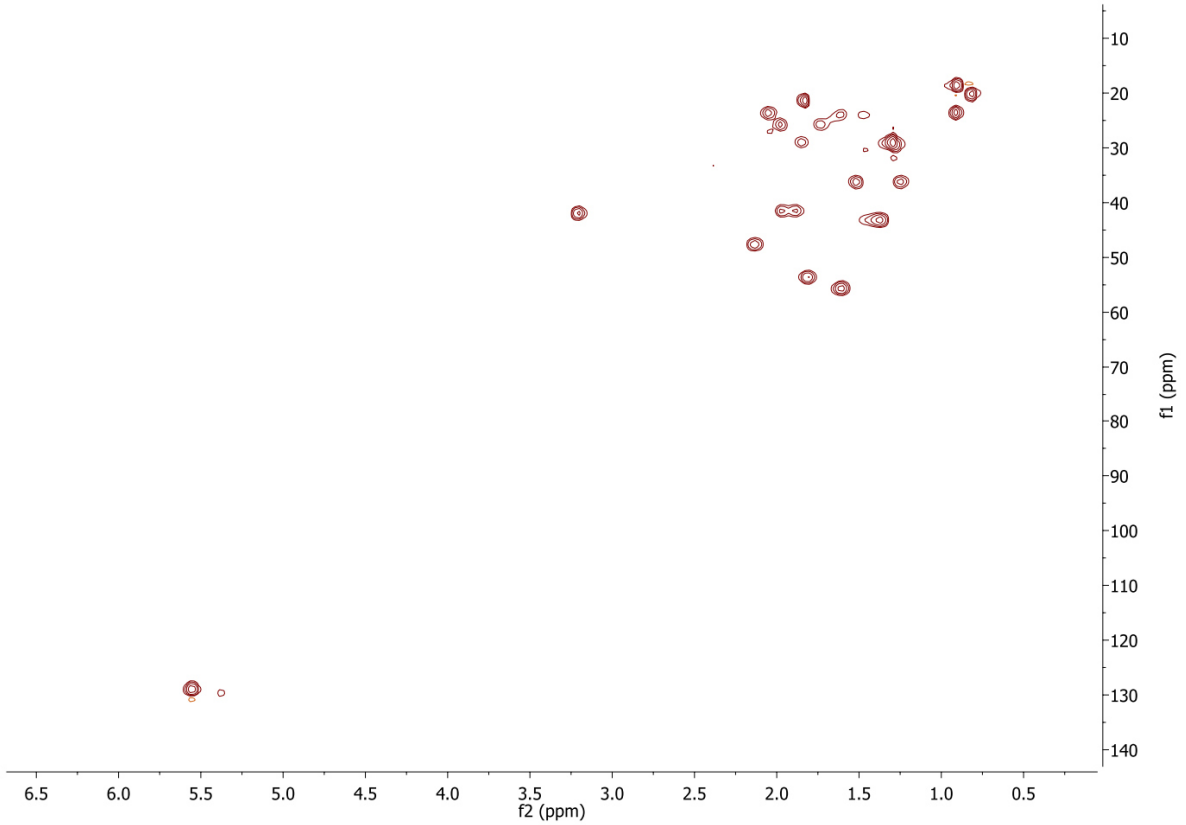


Figure 35. HMBC NMR (500 MHz, CDCl3) of cyclooctat-7-en-3-ol

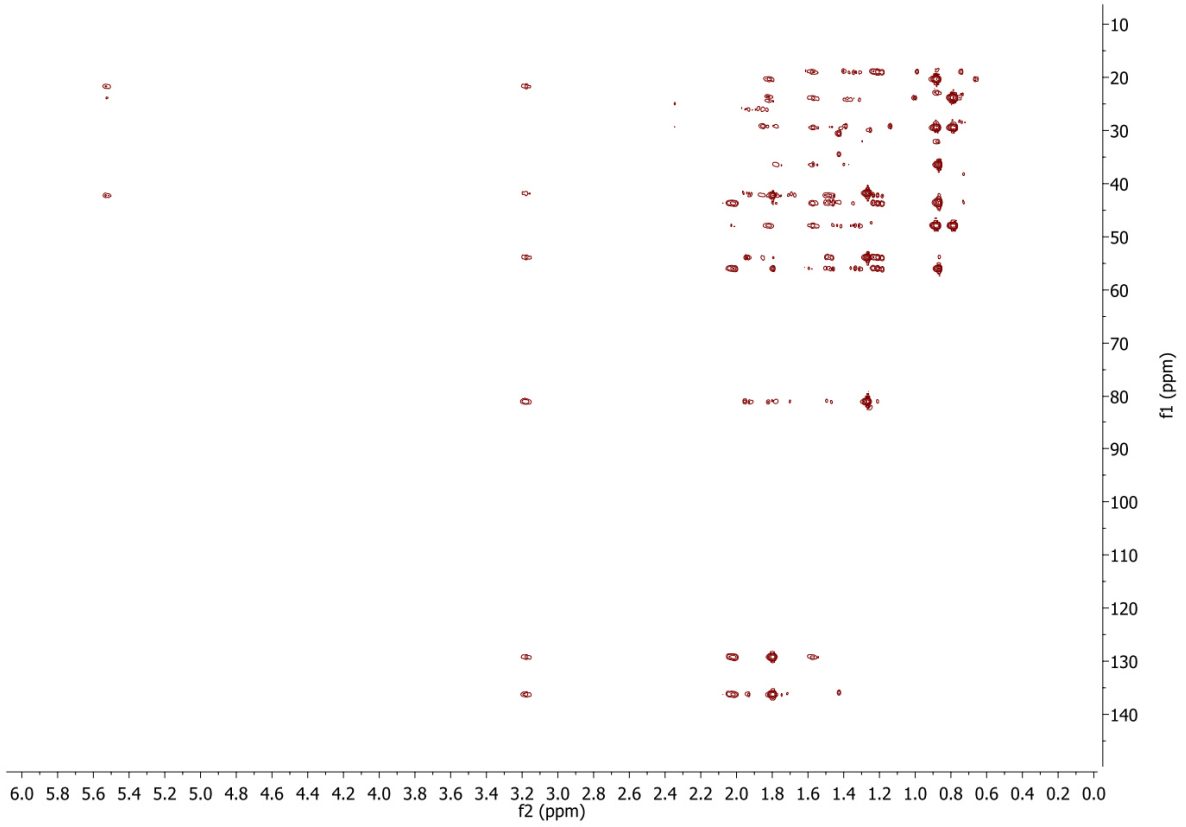


Figure 36. NOESY NMR (500 MHz, CDCl3) of cyclooctat-7-en-3-ol

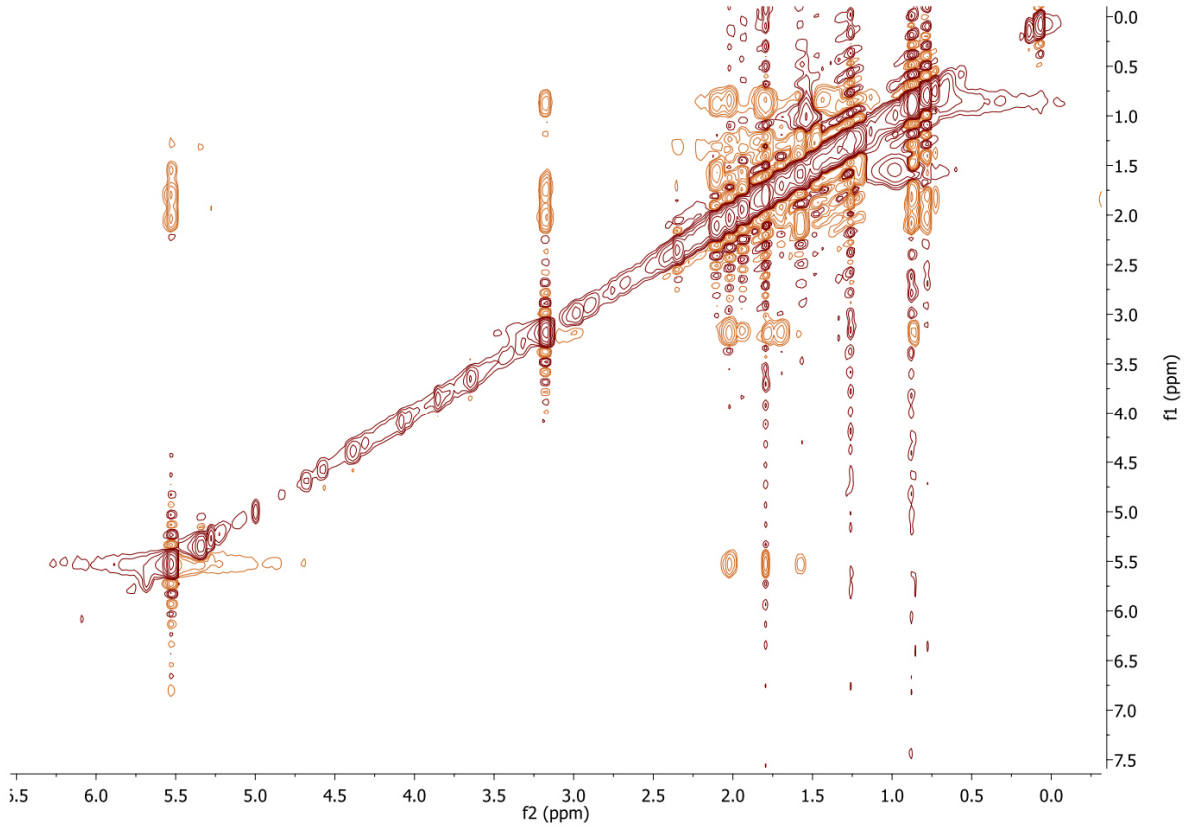


Figure 37. [1]H NMR (500 MHz, CDCl3) of cyclooctat-1,7-diene

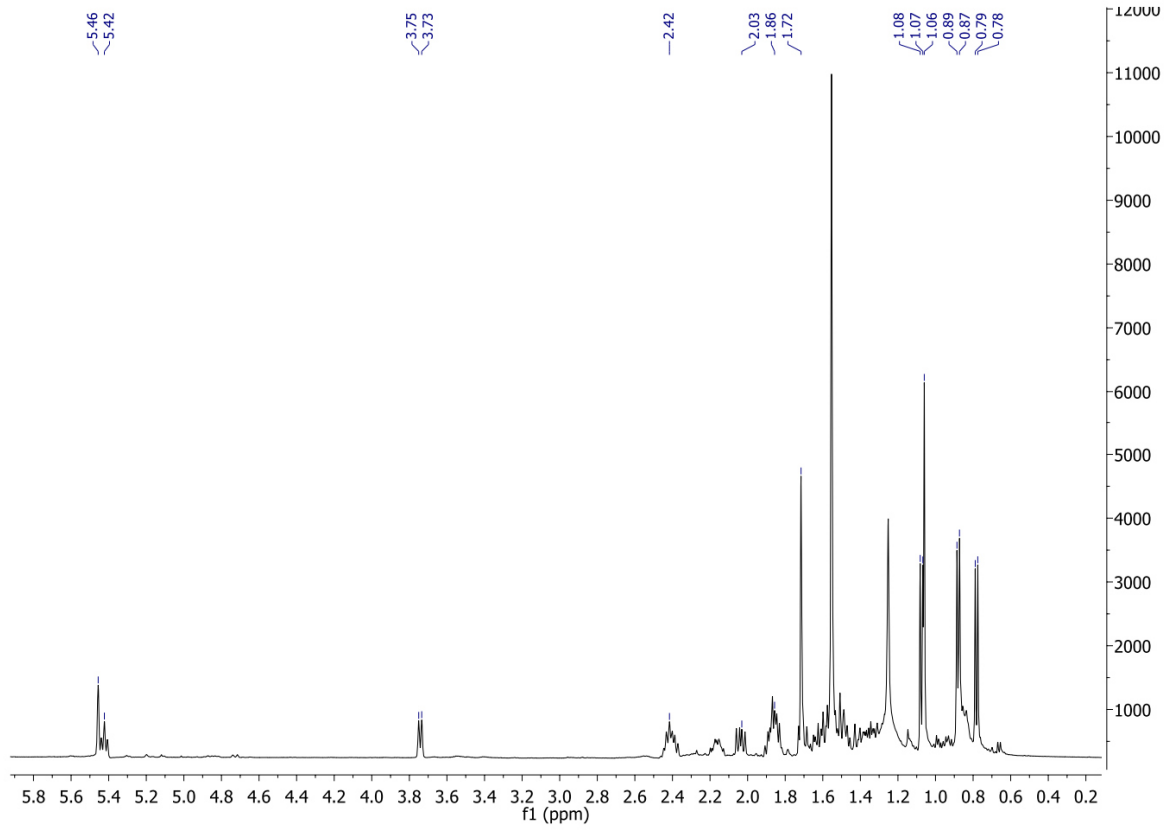


Figure 38. ^{13}C NMR (125 MHz, $CDCl_3$) of cyclooctat-1,7-diene

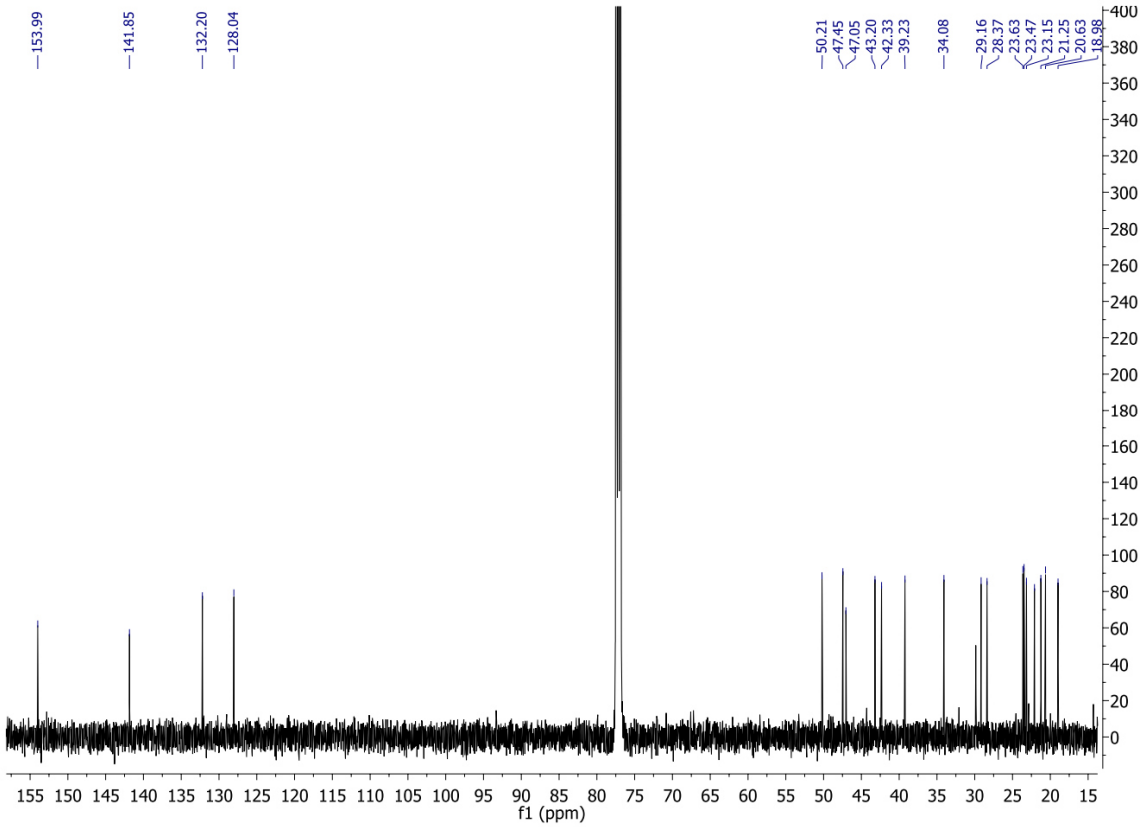


Figure 39. COSY NMR (500 MHz, CDCl3) of cyclooctat-1,7-diene

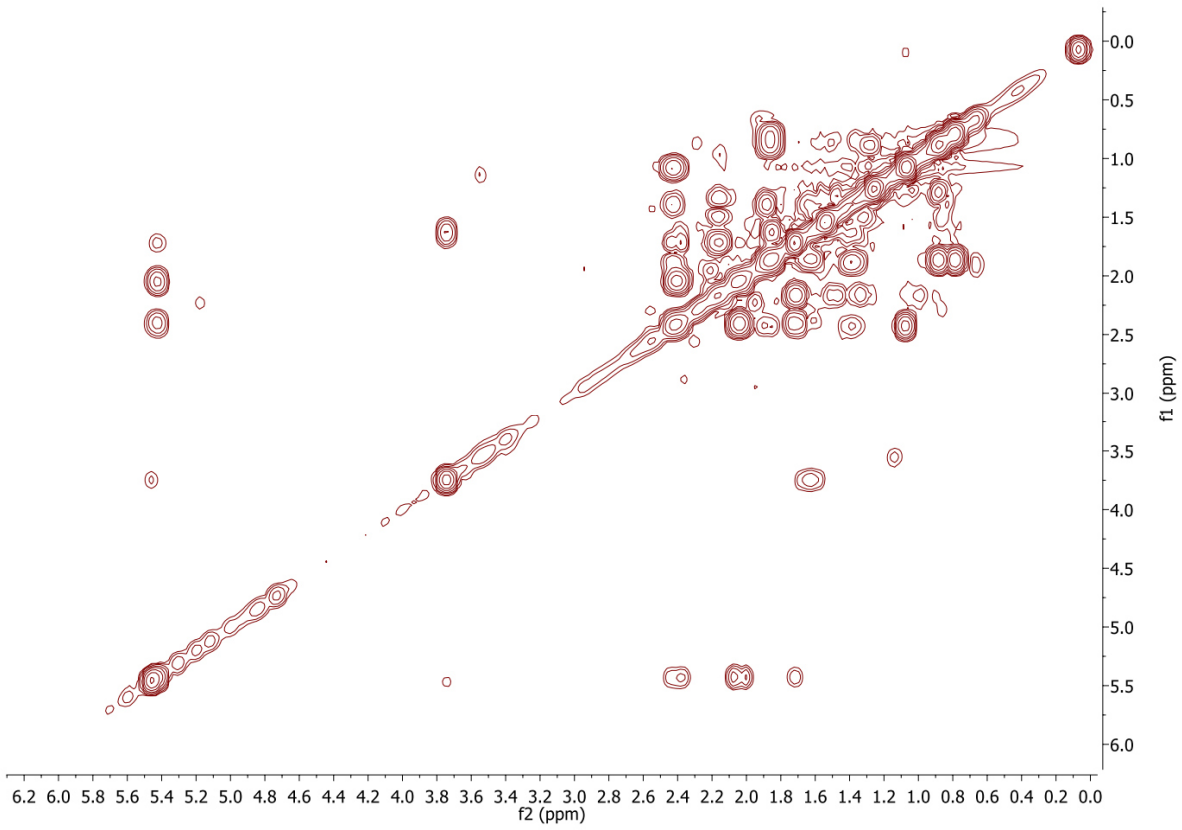


Figure 40. HSQC NMR (500 MHz, CDCl3) of cyclooctat-1,7-diene

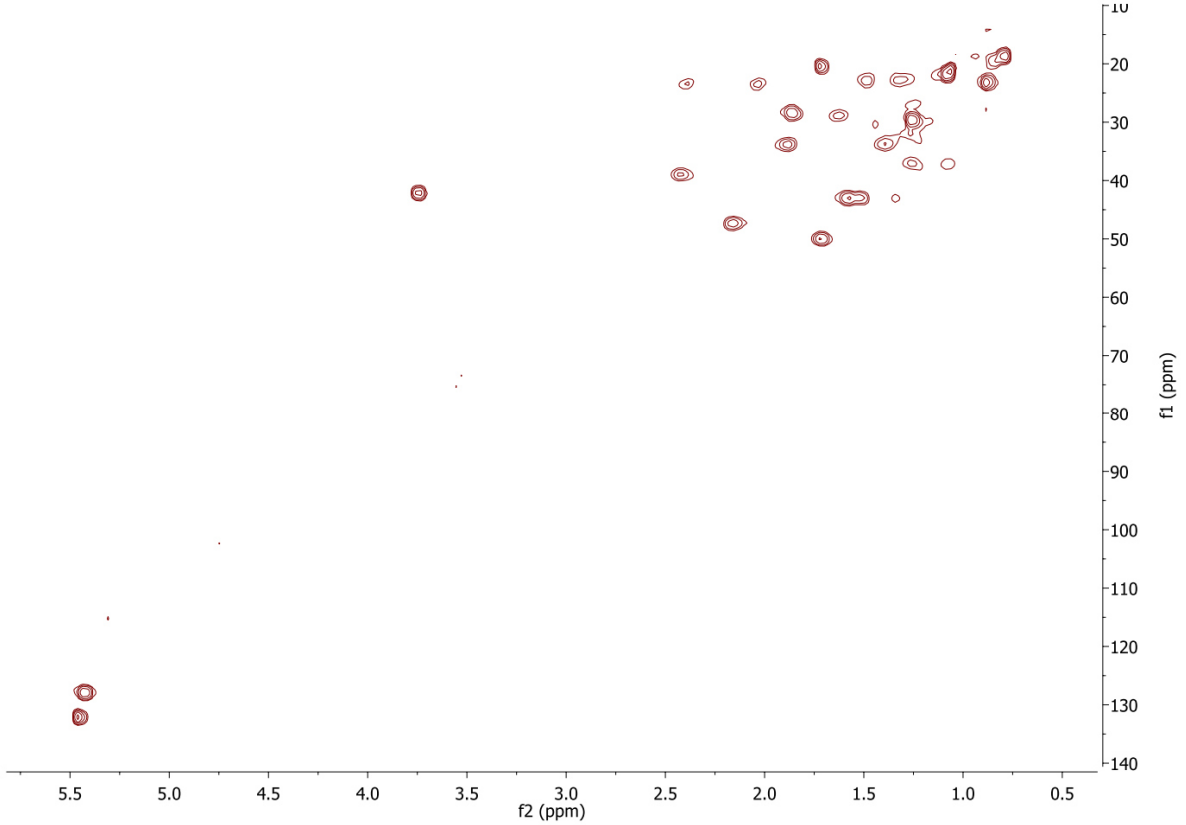


Figure 41. HMBC NMR (500 MHz, CDCl3) of cyclooctat-1,7-diene

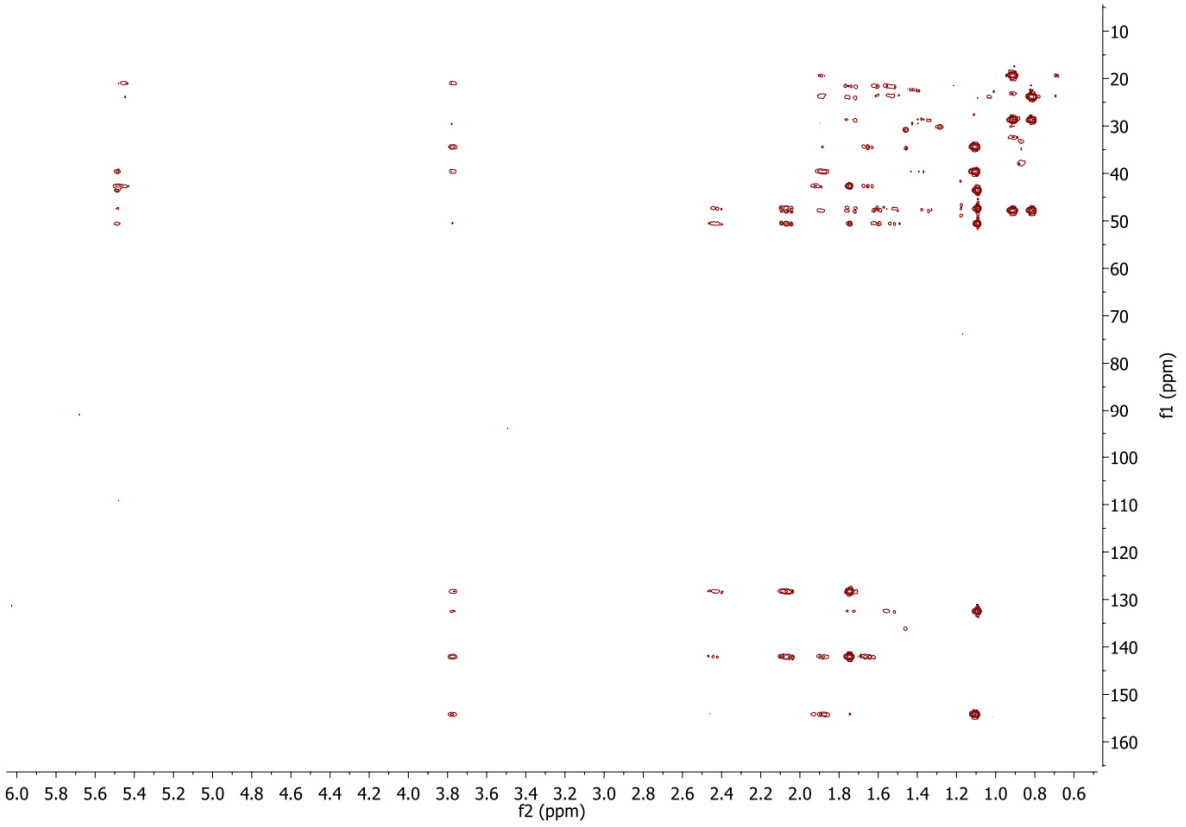


Figure 42. NOESY NMR (500 MHz, CDCl3) of cyclooctat-1,7-diene

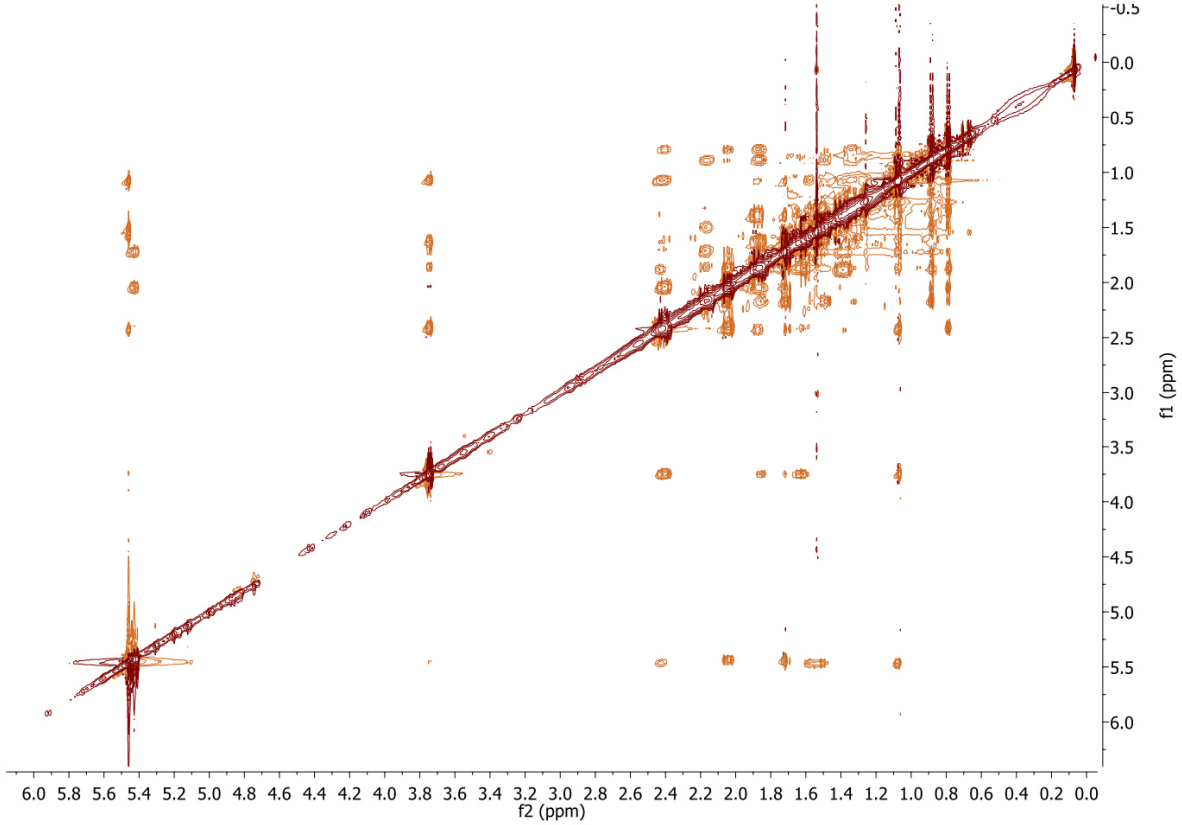


Figure 43. ^{1}H NMR (500 MHz, CDCl3) of (1R,3E,7E,11S,12S)-3,7,18-dolabellatriene

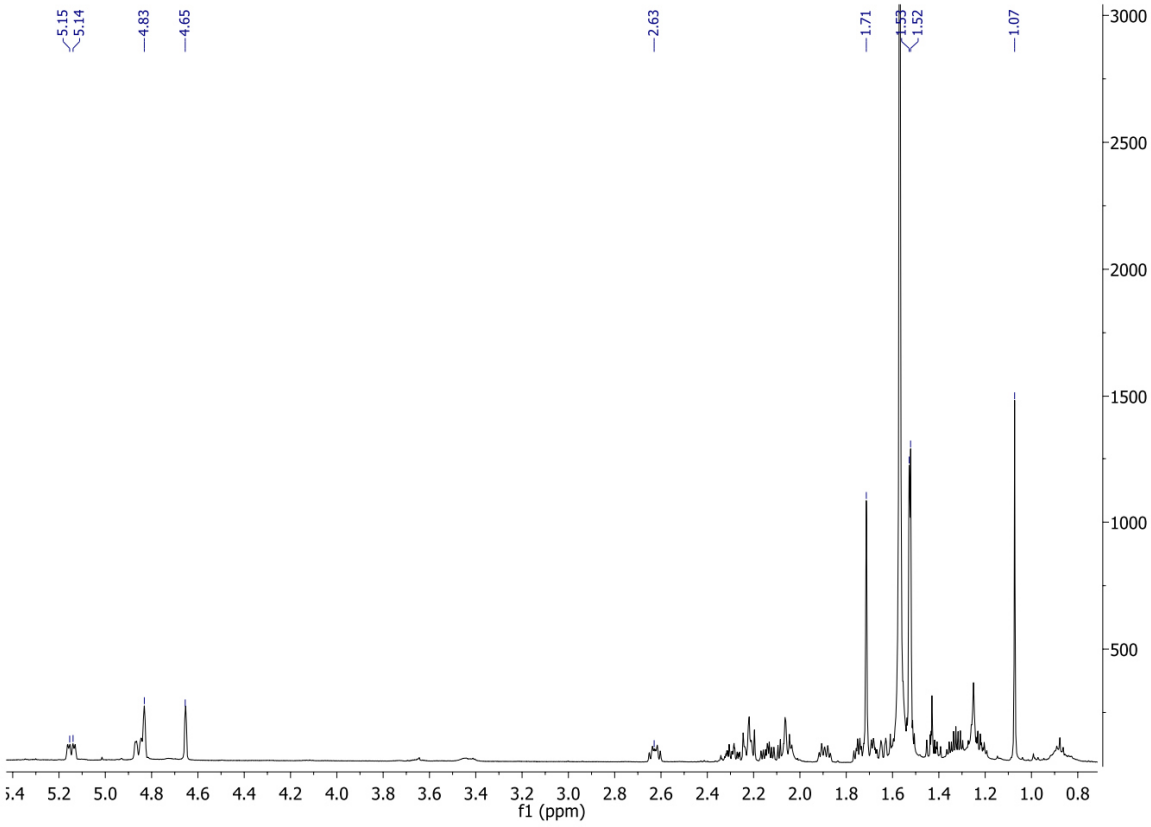


Figure 44. 13C NMR (125 MHz, CDCl3) of (1R,3E,7E,11S,12S)-3,7,18-dolabellatriene

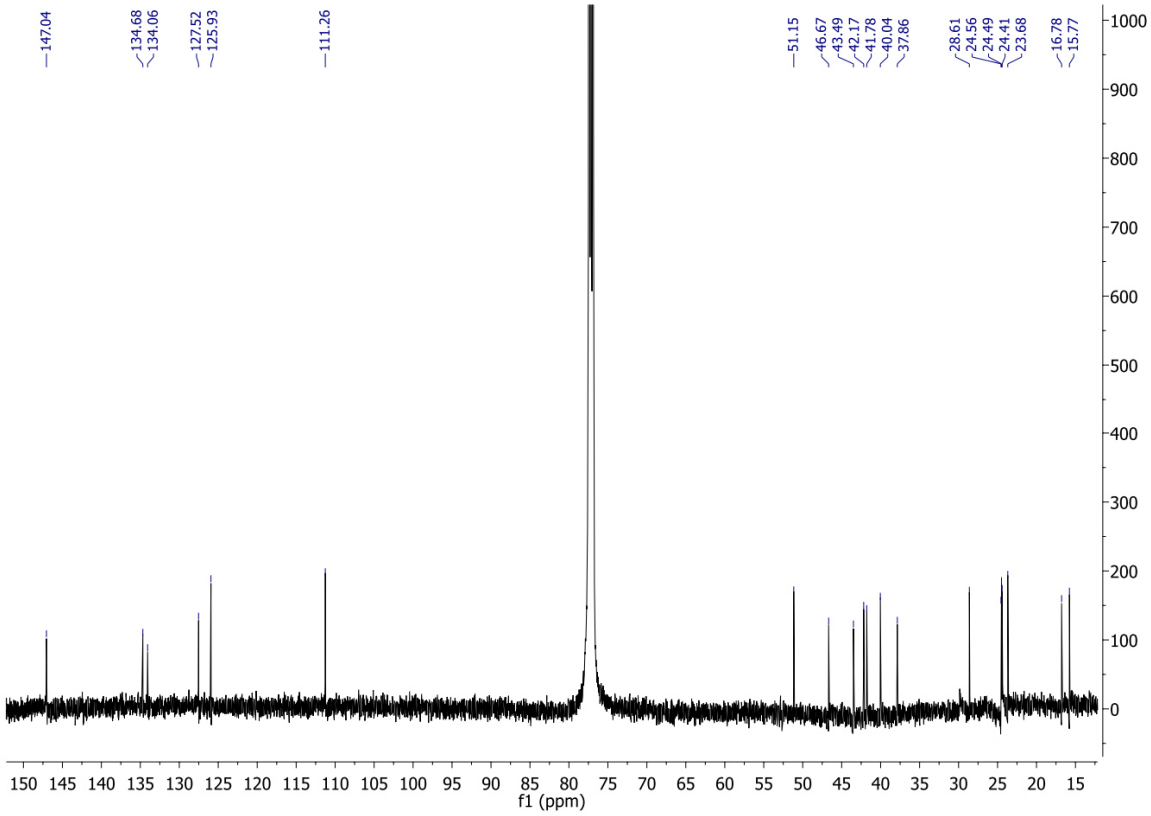


Figure 45. COSY NMR (500 MHz, CDCl3) of (1R,3E,7E,11S,12S)-3,7,18-dolabellatriene

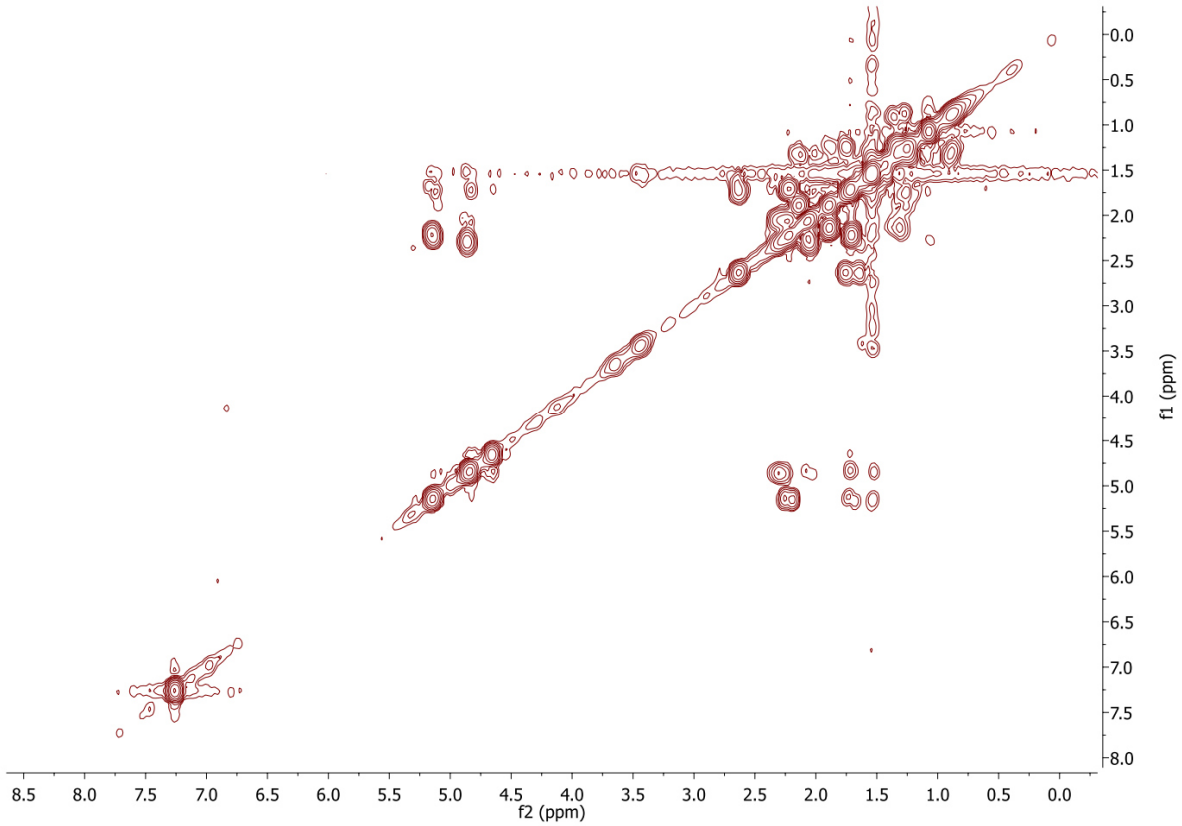


Figure 46. HSQC NMR (500 MHz, CDCl3) of (1R,3E,7E,11S,12S)-3,7,18-dolabellatriene

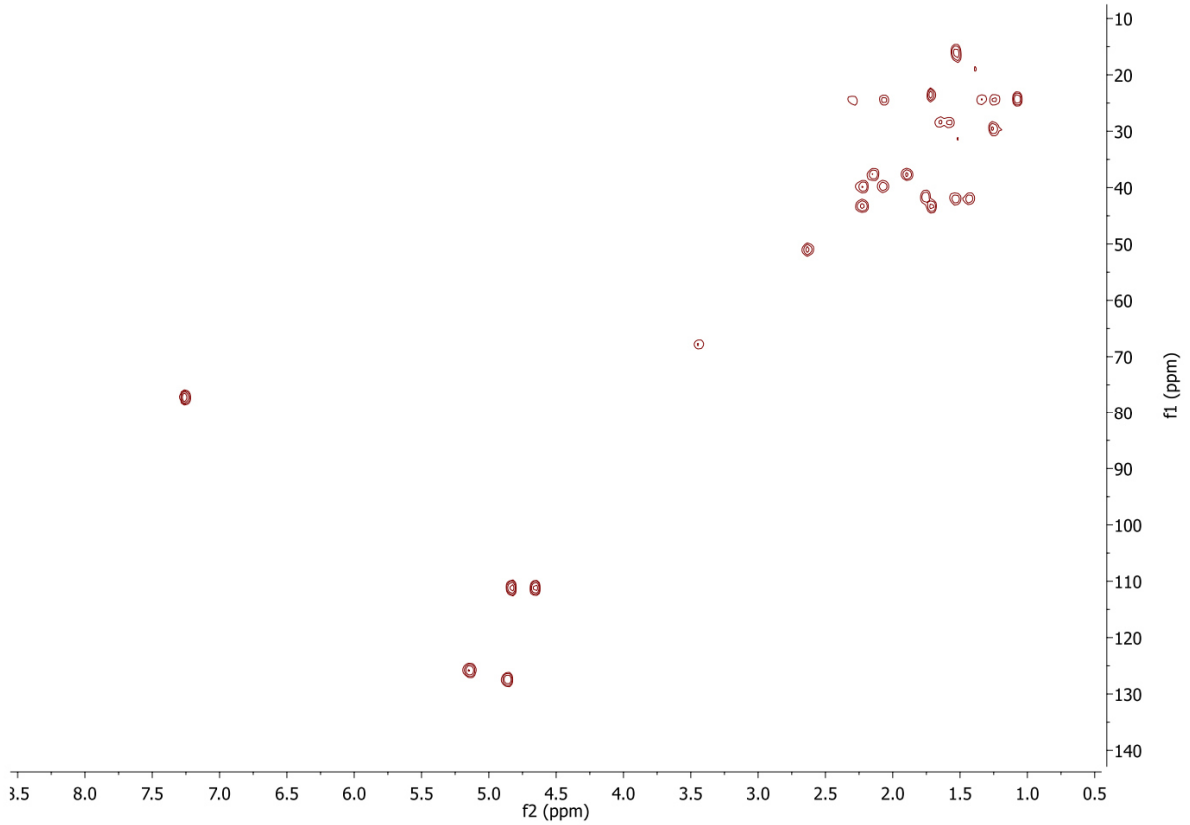


Figure 47. HMBC NMR (500 MHz, CDCl3) of (1R,3E,7E,11S,12S)-3,7,18-dolabellatriene

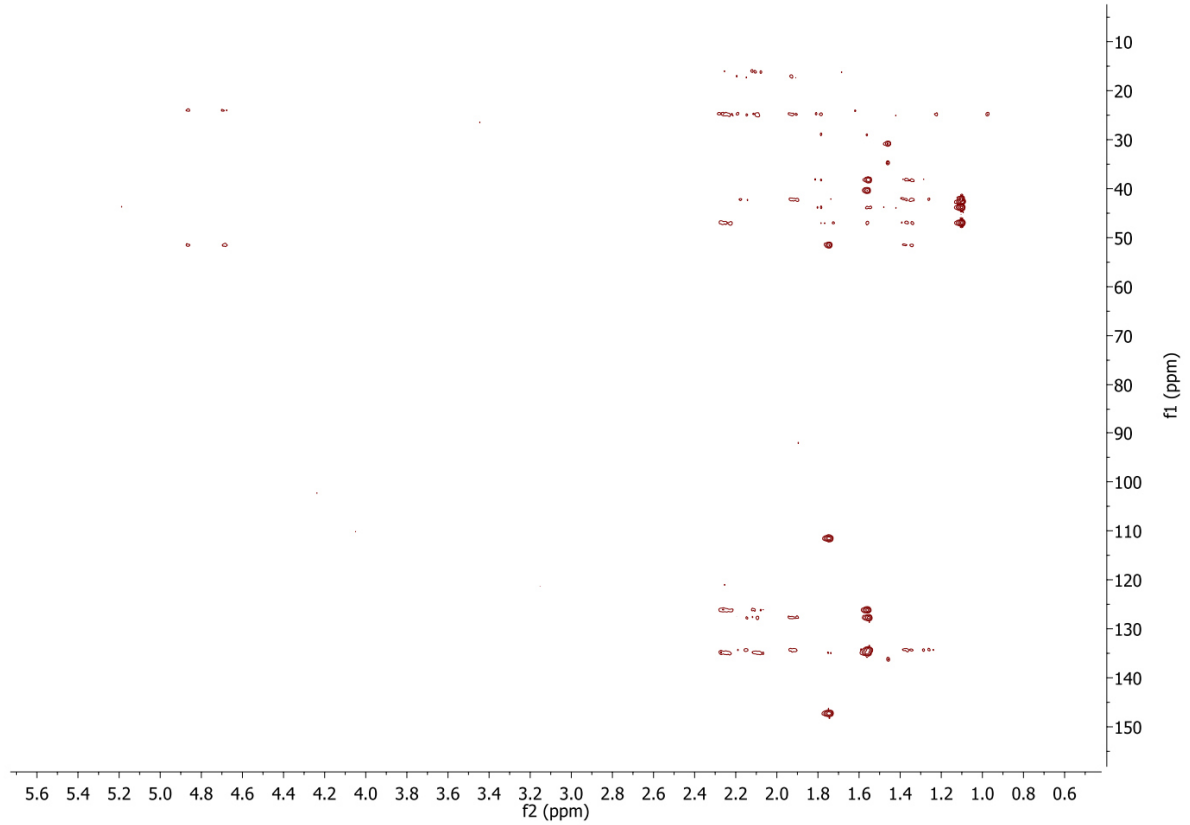


Figure 48. NOESY NMR (500 MHz, CDCl3) of (1R,3E,7E,11S,12S)-3,7,18-dolabellatriene

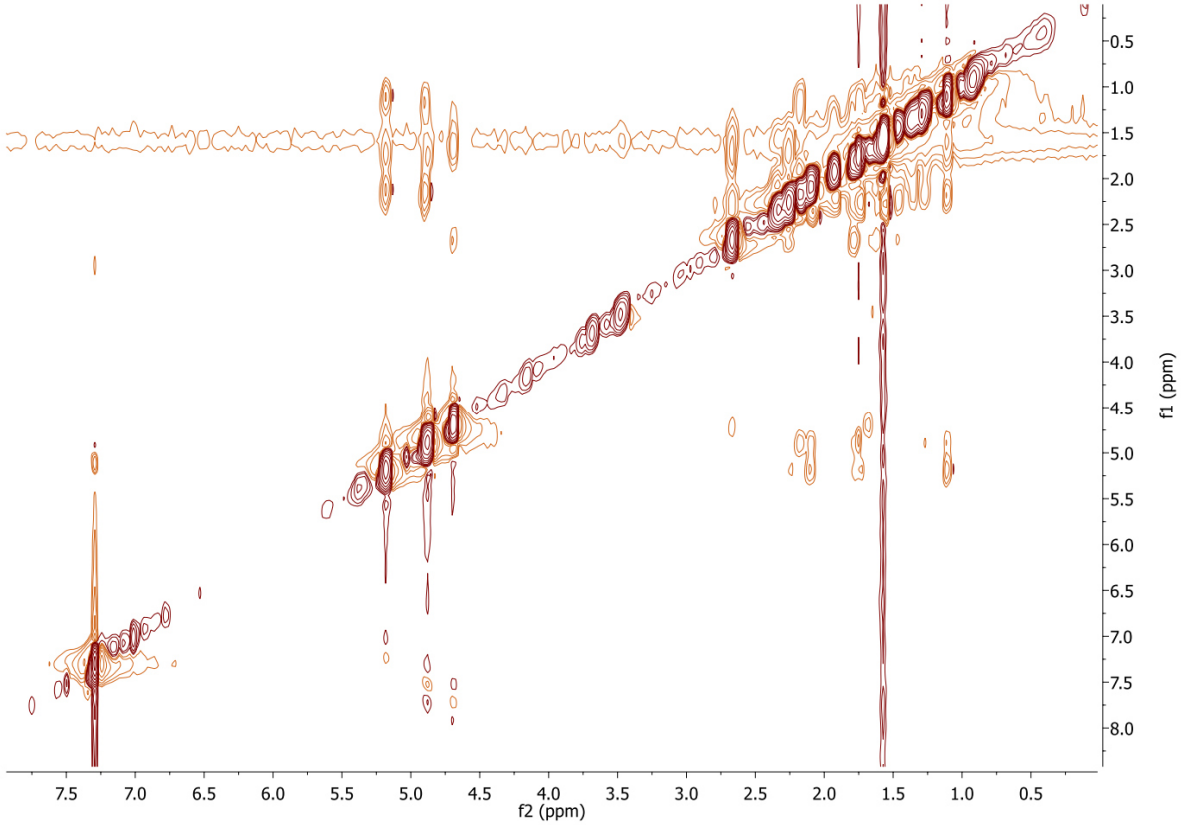


Figure 49. ^{1}H NMR (500 MHz, CDCl3) of cyclooctat-9-en-5,7-diol

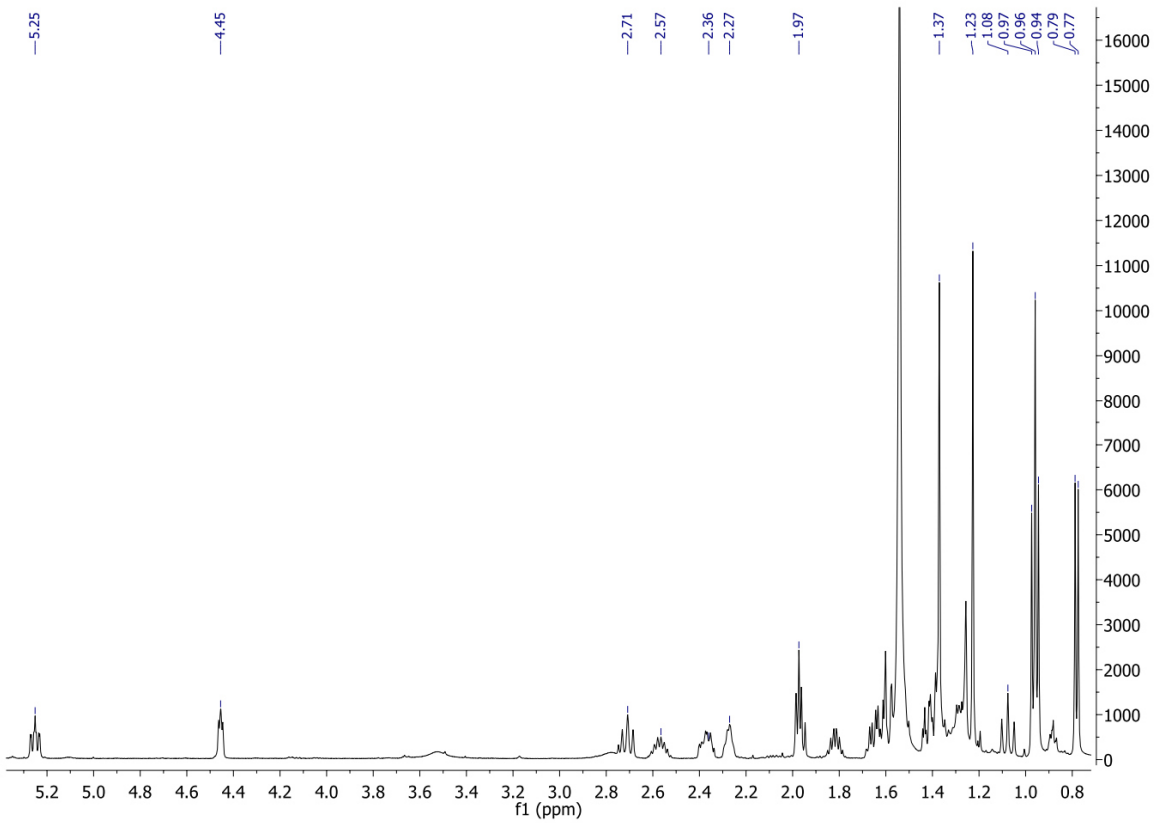


Figure 50. ^{13}C NMR (125 MHz, $CDCl_3$) of cyclooctat-9-en-5,7-diol

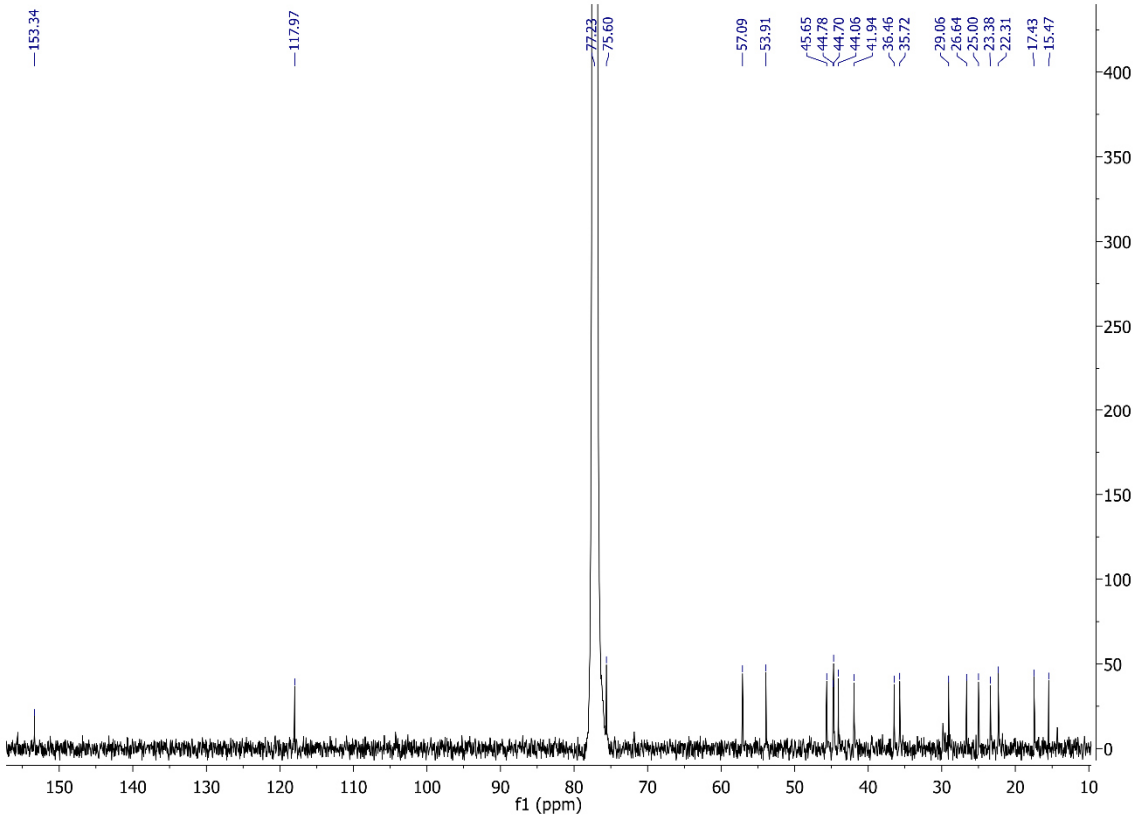


Figure 51. ^{1}H NMR (500 MHz, CD_3-OD) of cyclooctatin

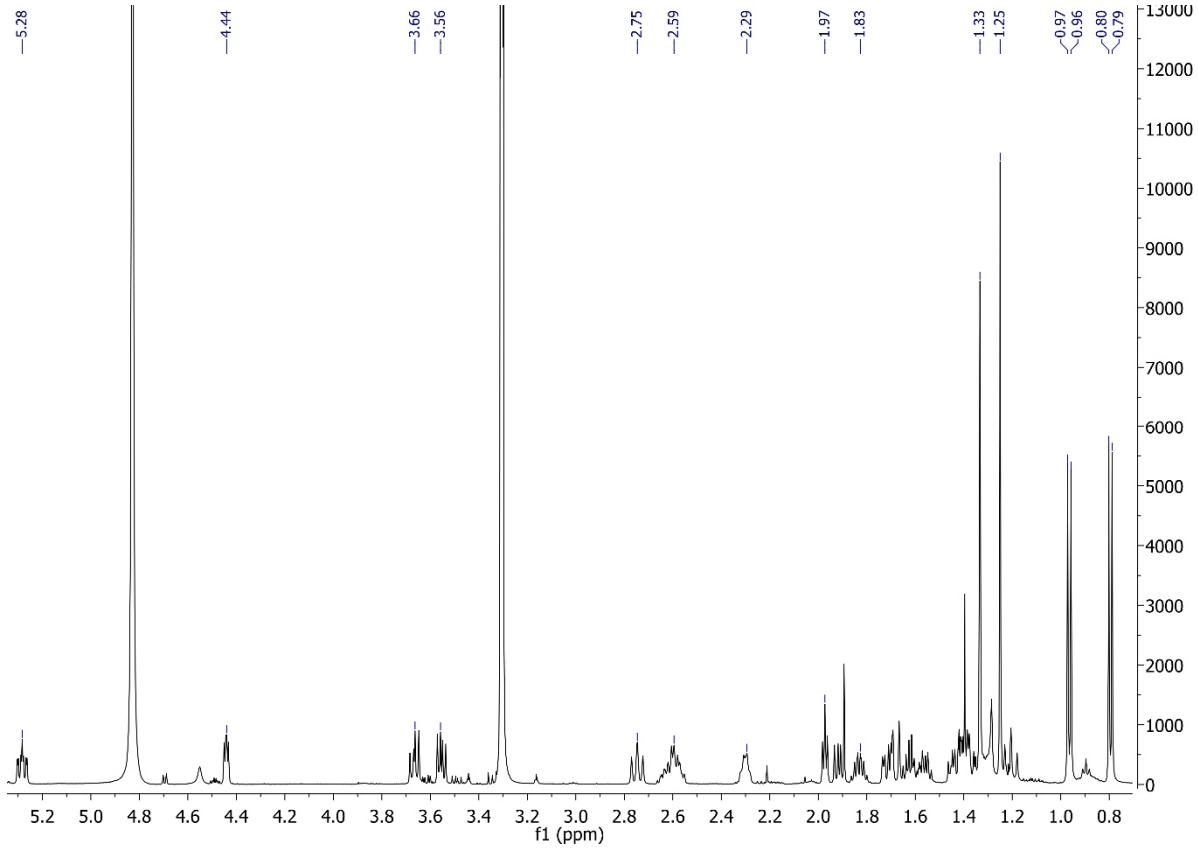


Figure 52. ^{13}C NMR (125 MHz, CD_3-OD) of cyclooctatin

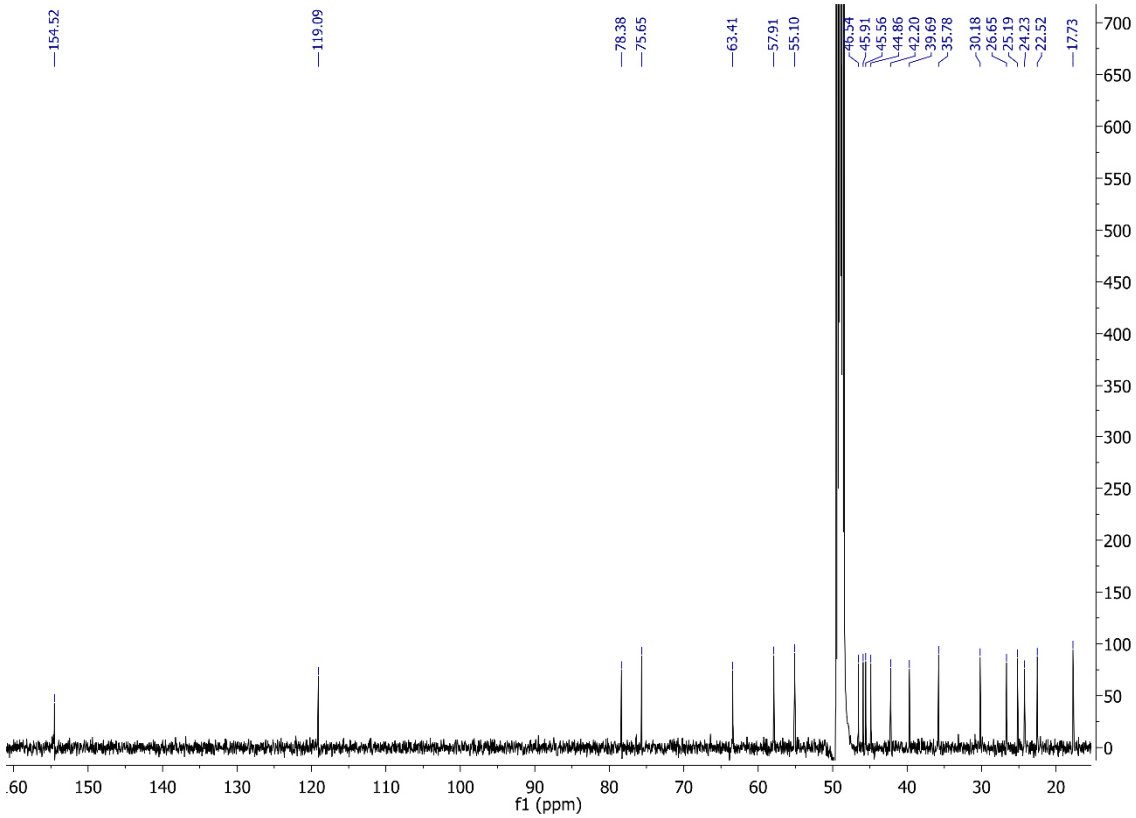


Figure 53. COSY NMR (500 MHz, CD_3-OD) of cyclooctatin

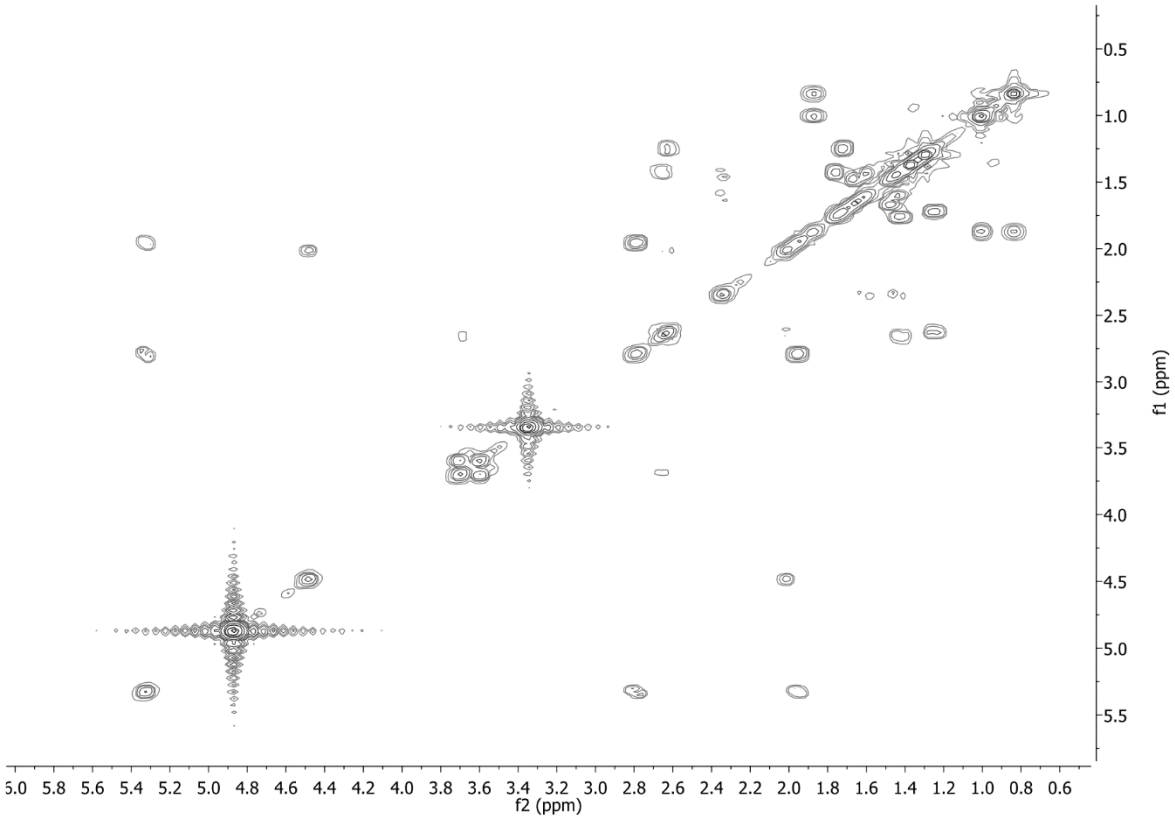


Figure 54. HSQC (500 MHz, CD_3-OD) NMR of cyclooctatin

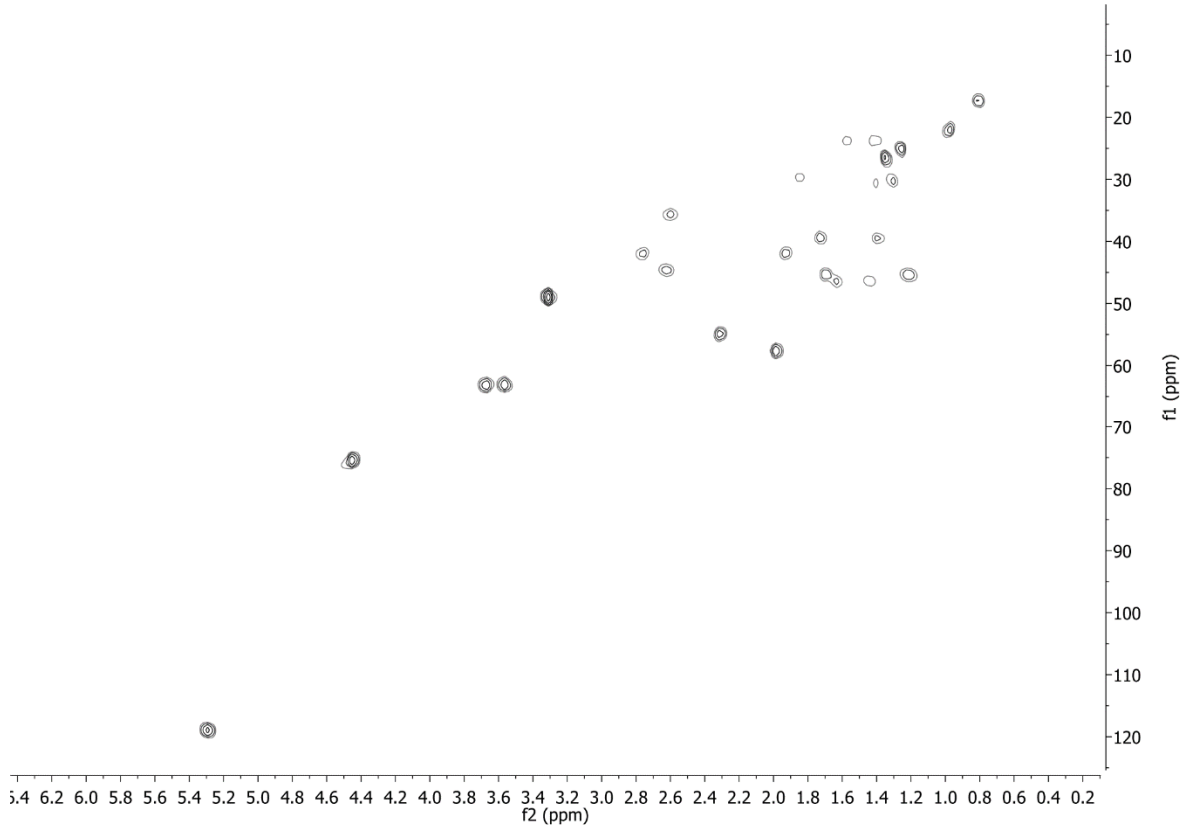


Figure 55. NOESY NMR (500 MHz, CD3-OD) of cyclooctatin

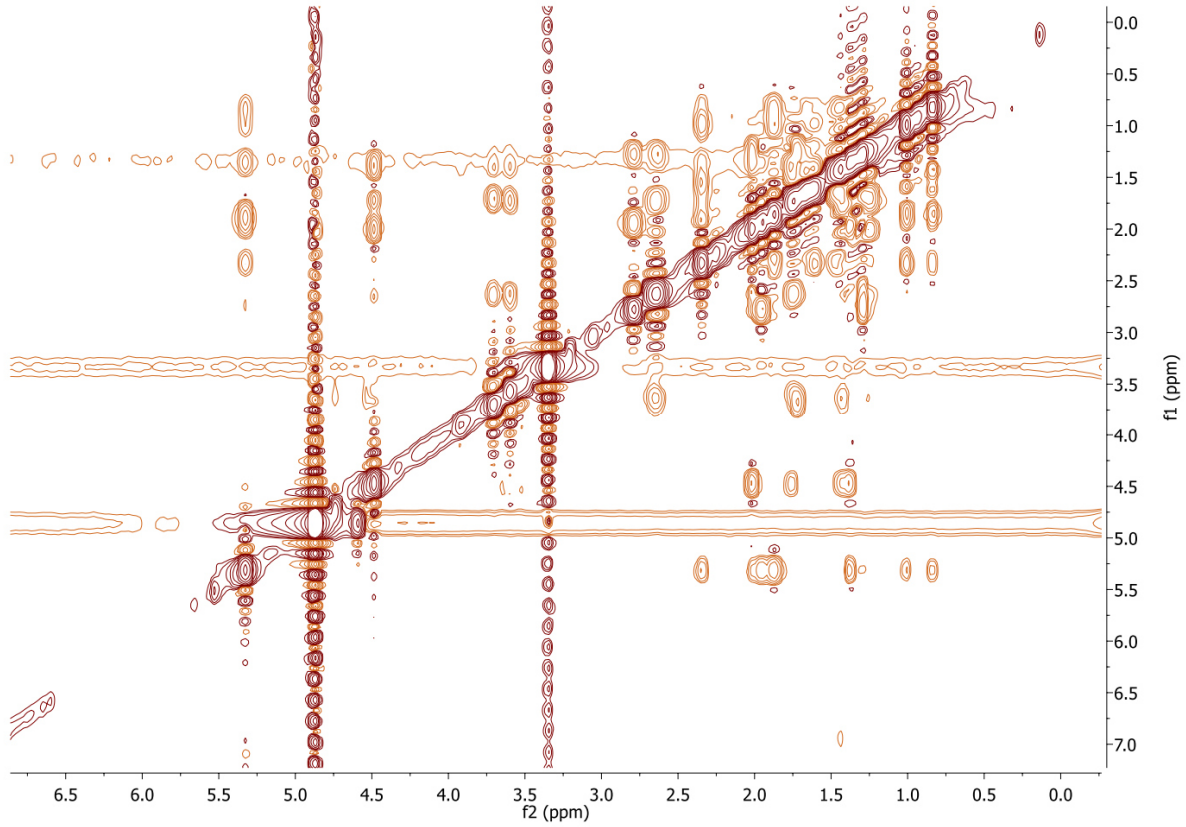


Figure 56. ^{1}H NMR (500 MHz, $CDCl_3$) of the C-5 R-MTPA derivative of cyclooctat-9-en-5,7-diol

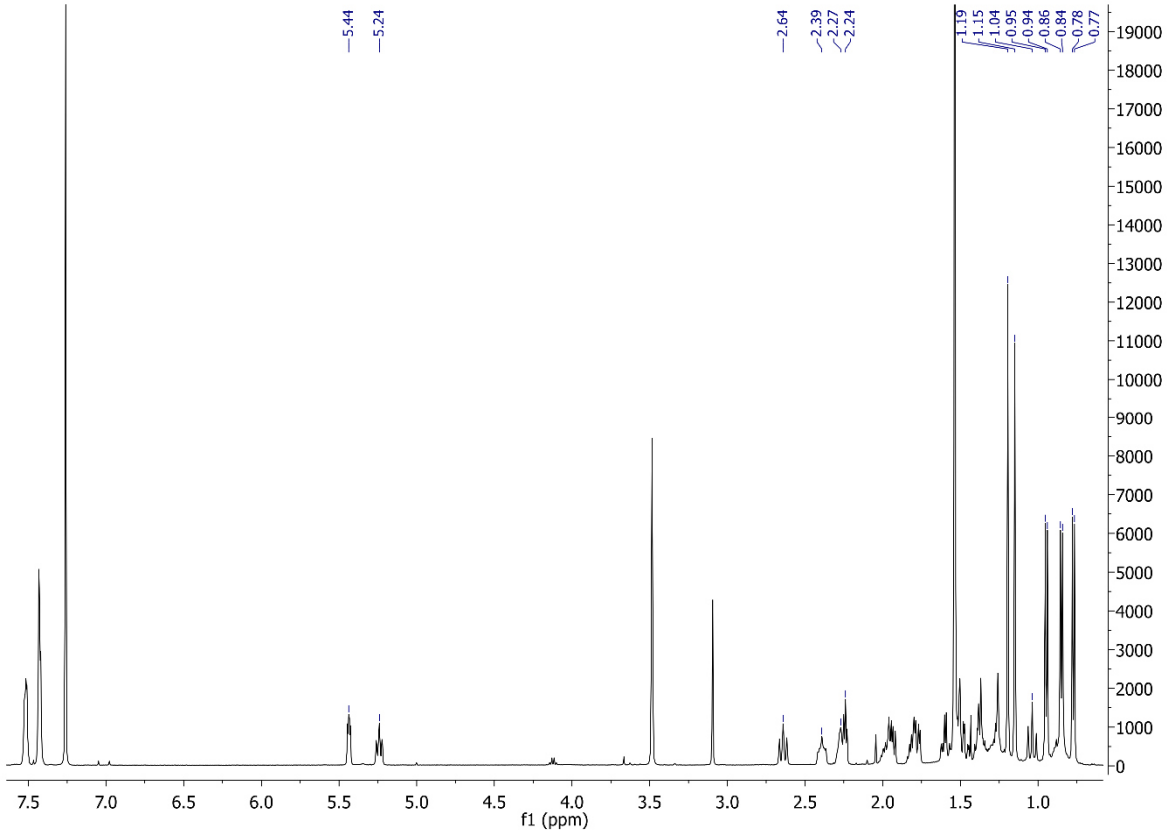


Figure 57. ^{13}C NMR (125 MHz, $CDCl_3$) of the C-5 R-MTPA derivative of cyclooctat-9-en-5,7-diol

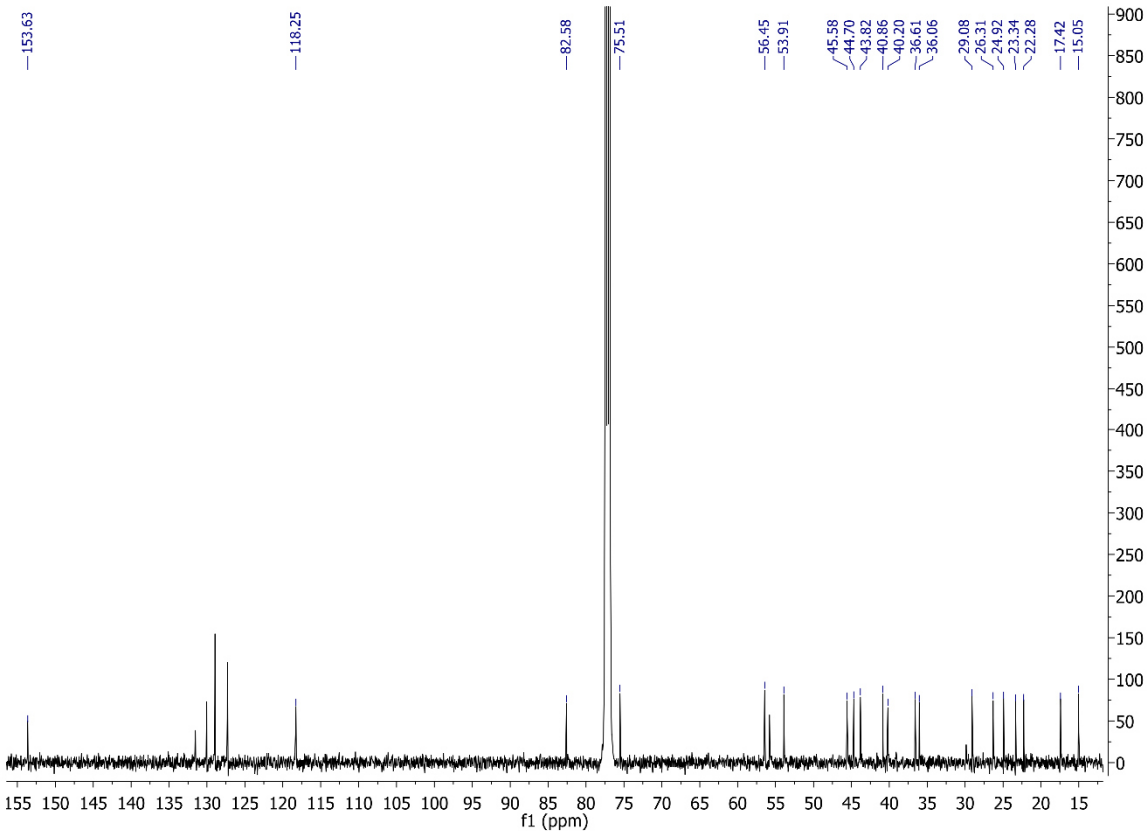


Figure 58. COSY NMR (500 MHz, $CDCl_3$) of the C-5 R-MTPA derivative of cyclooctat-9-en-5,7-diol

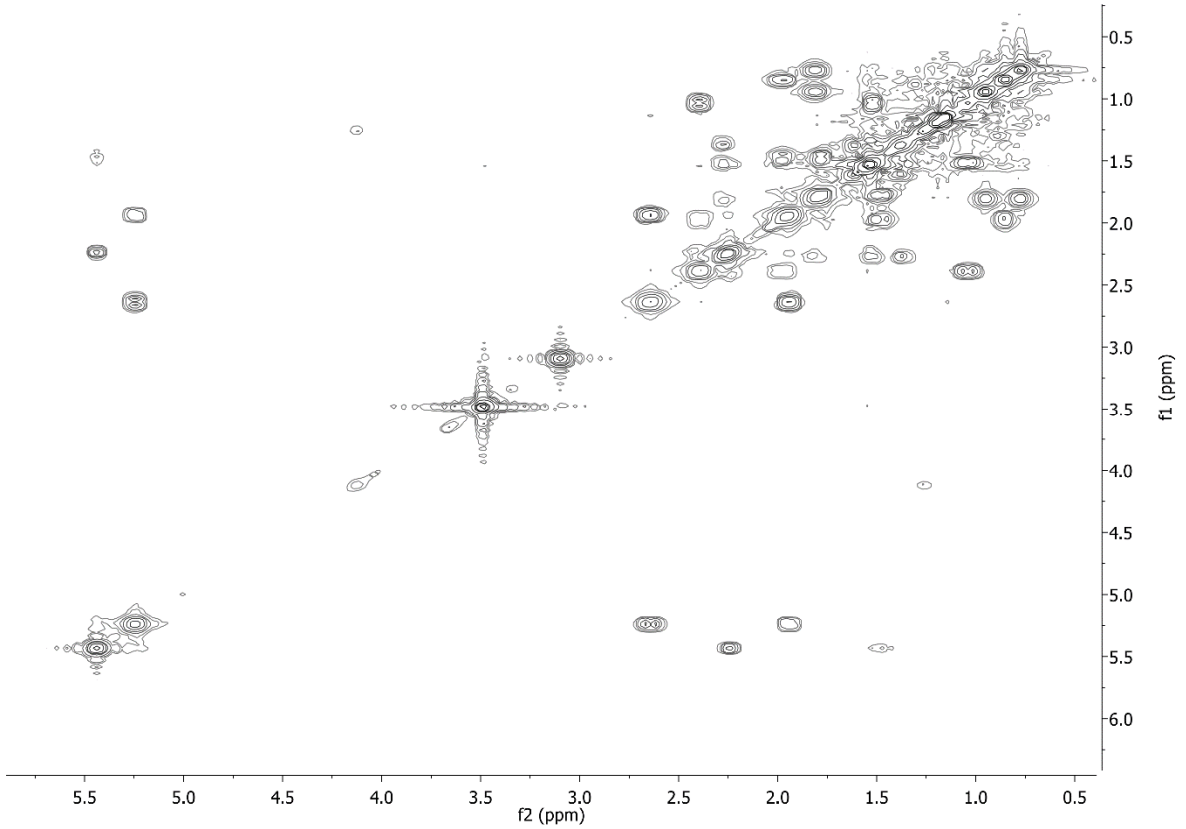


Figure 59. HSQC NMR (500 MHz, $CDCl_3$) of the C-5 R-MTPA derivative of cyclooctat-9-en-5,7-diol

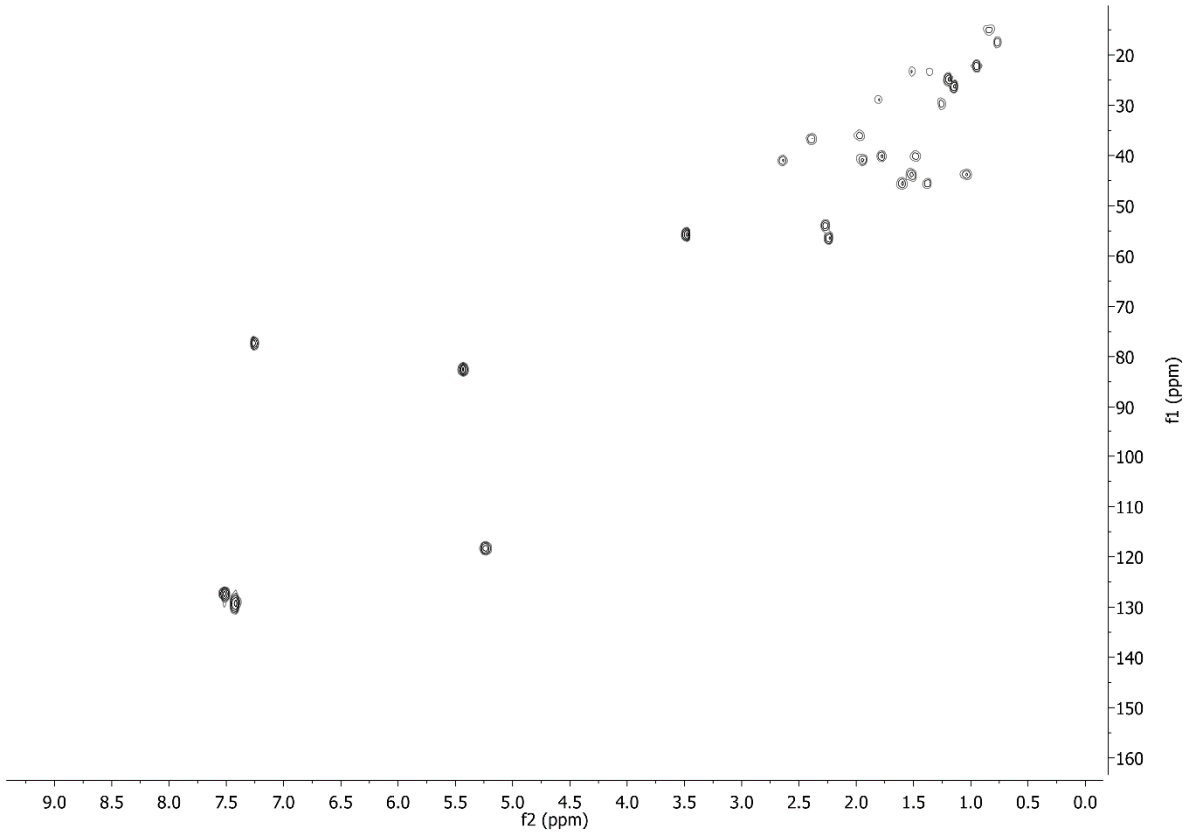


Figure 60. ^{1}H NMR (500 MHz, $CDCl_3$) of the C-5 S-MTPA derivative of cyclooctat-9-en-5,7-diol

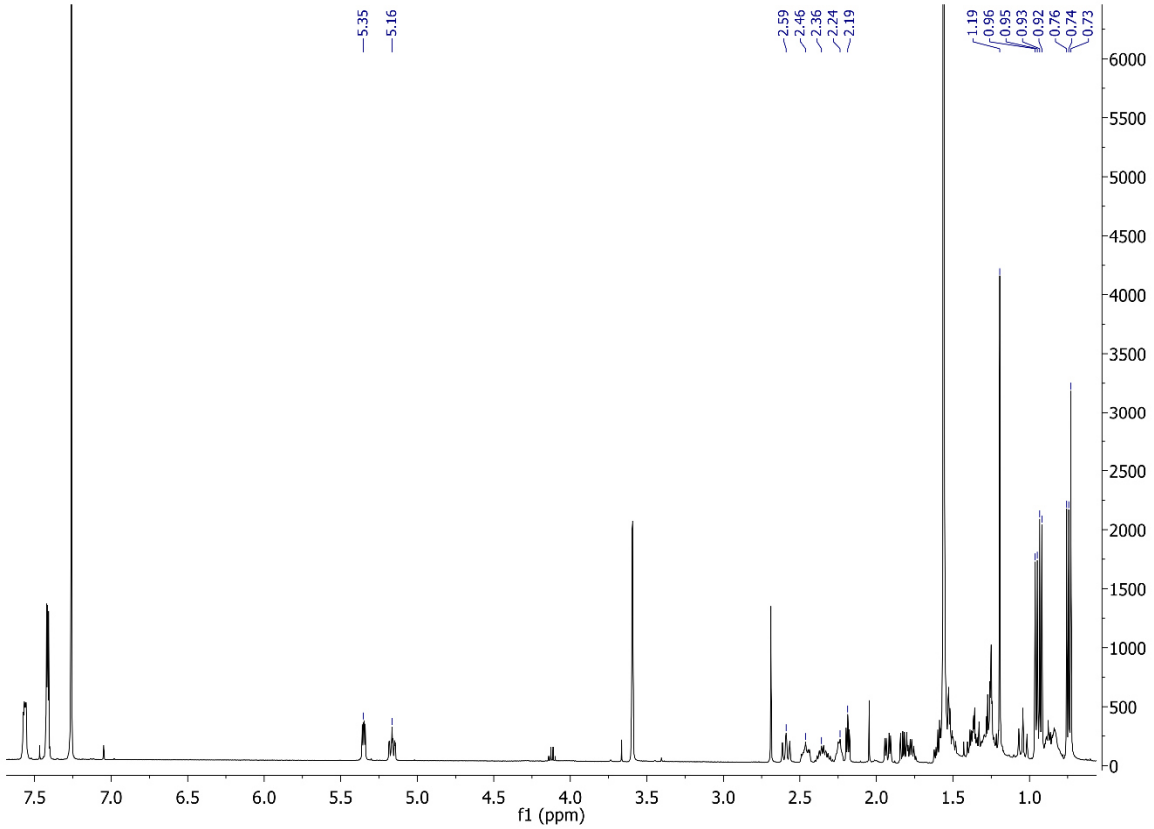


Figure 61. ^{13}C NMR (125 MHz, $CDCl_3$) of the C-5 S-MTPA derivative of cyclooctat-9-en-5,7-diol

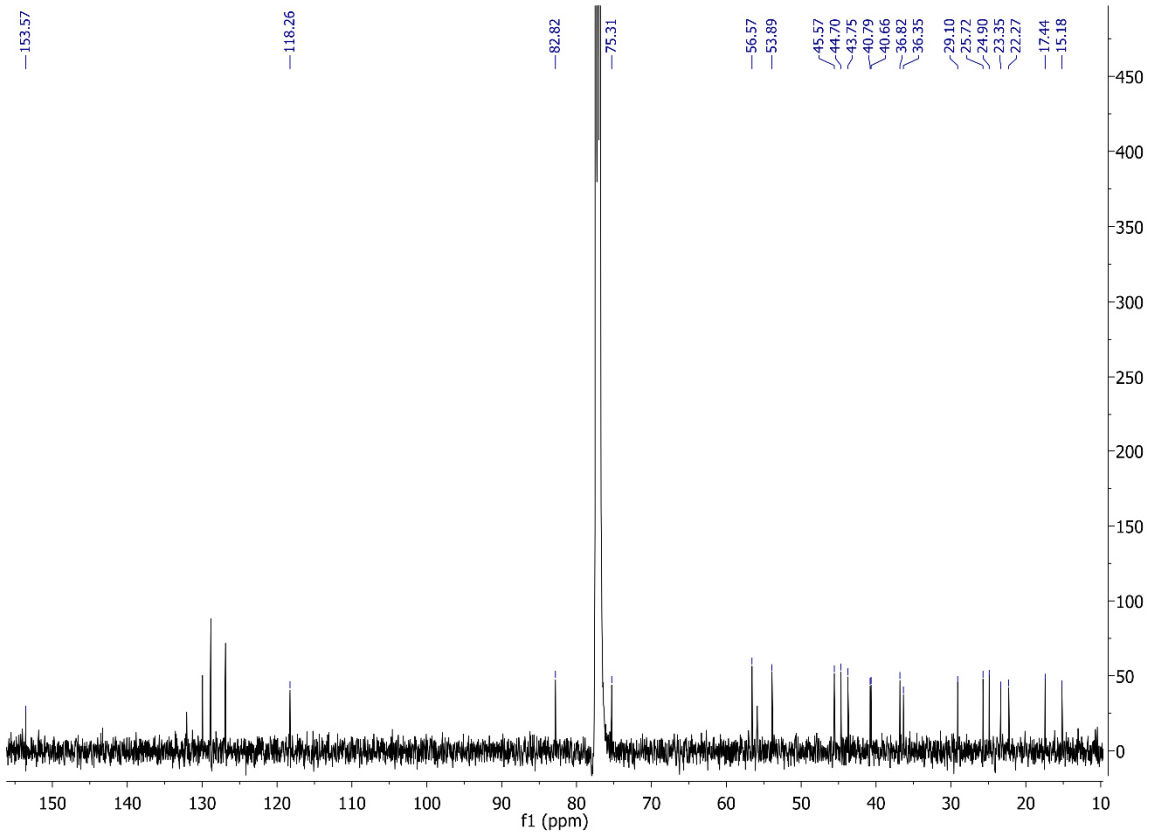


Figure 62. COSY NMR (500 MHz, $CDCl_3$) of the C-5 S-MTPA derivative of cyclooctat-9-en-5,7-diol

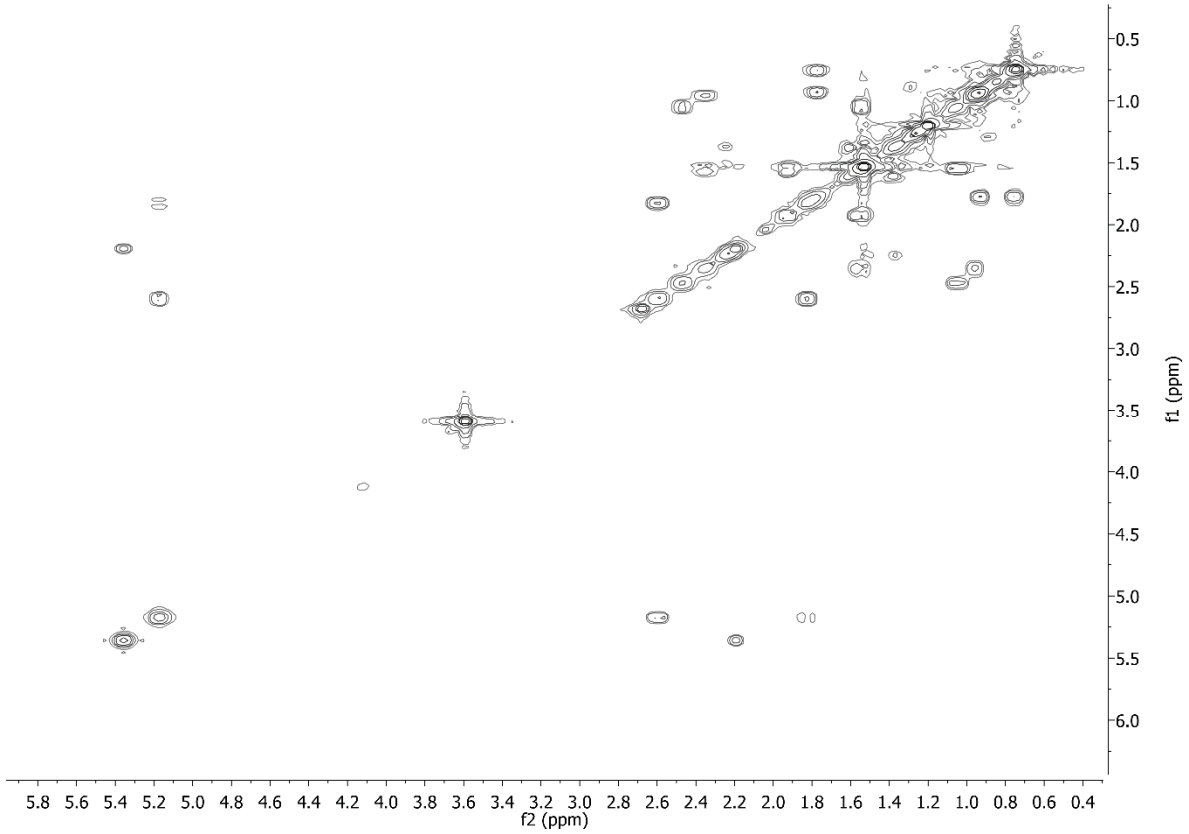


Figure 63. HSQC NMR (500 MHz, $CDCl_3$) of the C-5 S-MTPA derivative of cyclooctat-9-en-5,7-diol

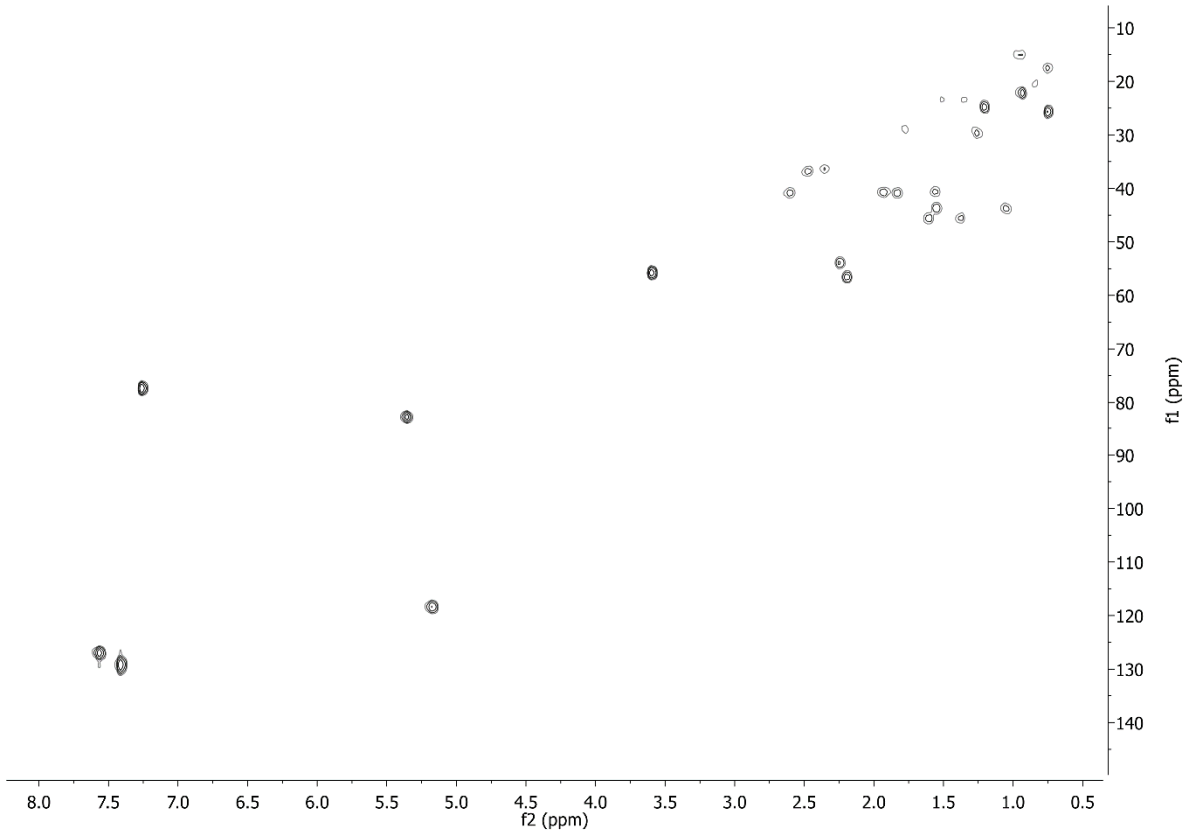


Figure 64. ^{1}H NMR (500 MHz, $CDCl_3$) Overlay of the C-5 R/S-MTPA derivatives of cyclooctat-9-en-5,7-diol

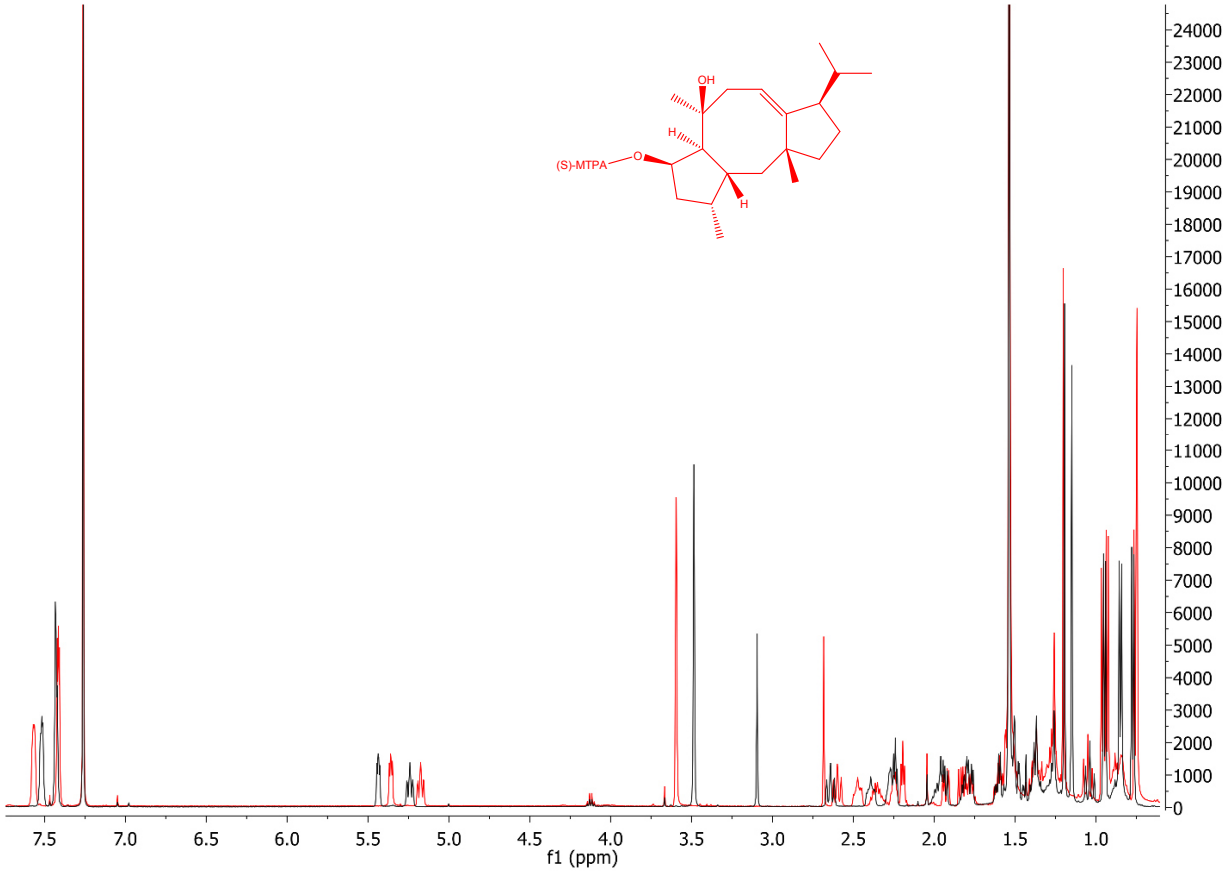
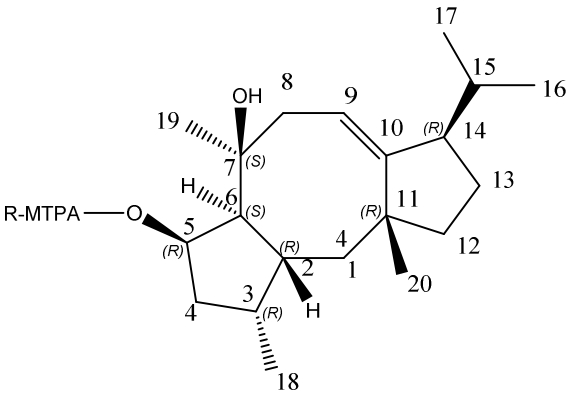
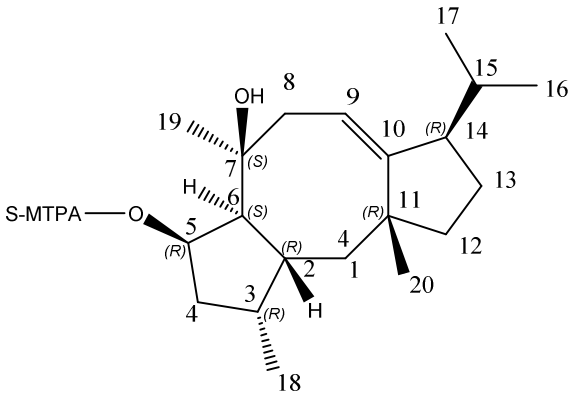


Table 3. ^{1}H NMR ($CDCl_3$, 500 MHz, δ_H) and ^{13}C ($CDCl_3$, 125 MHz, δ_C) data of the C-5 R-MTPA derivatives of cyclooctat-9-en-5,7-diol



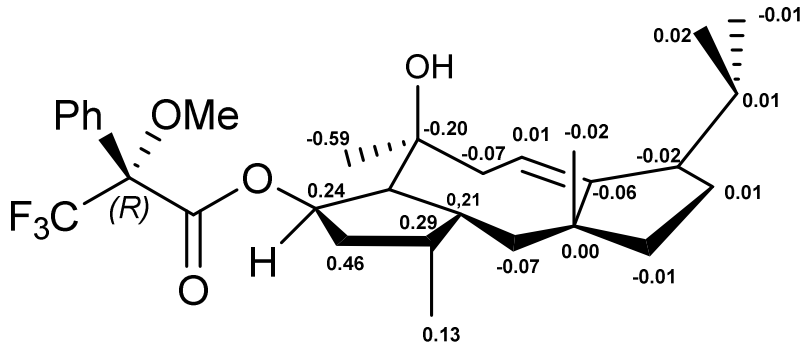
Position	δ_C	δ_H
1	43.82	1.04, 1.52
2	36.61	2.39
3	36.06	1.97
4	40.20	1.49, 1.78
5	82.58	5.43
6	56.45	2.24
7	75.51	
8	40.86	1.94, 2.64
9	118.25	5.24
10	153.63	
11	44.70	
12	45.58	1.38, 1.60
13	23.34	1.36, 1.51
14	53.91	2.27
15	29.09	1.26
16	17.42	0.77
17	22.28	0.95
18	15.05	0.84
19	26.31	1.14
20	24.92	1.19

Table 4. ^{1}H NMR ($CDCl_3$, 500 MHz, δ_H) and ^{13}C ($CDCl_3$, 125 MHz, δ_C) data of the C-5 S-MTPA derivatives of cyclooctat-9-en-5,7-diol

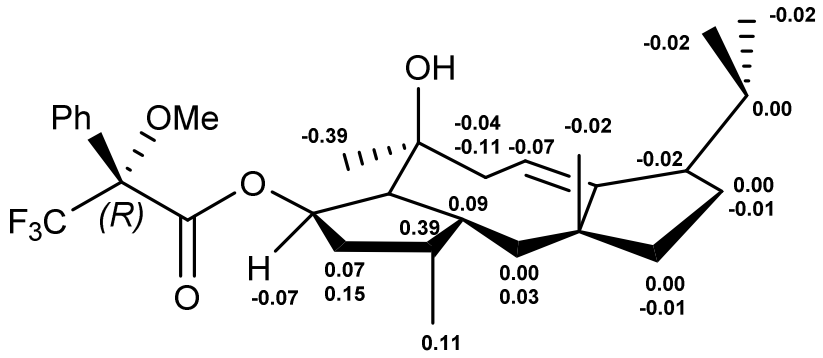


Position	δ_C	δ_H
1	43.75	1.04, 1.55
2	36.82	2.48
3	36.35	2.35
4	40.66	1.56, 1.93
5	82.82	5.36
6	56.57	2.19
7	75.31	
8	40.79	1.83, 2.60
9	118.26	5.17
10	153.57	
11	44.7	
12	45.57	1.37, 1.60
13	23.35	1.35, 1.51
14	53.89	2.25
15	29.10	1.26
16	17.44	0.75
17	22.27	0.93
18	15.18	0.95
19	25.72	0.75
20	24.9	1.21

Figure 65. Δδ (S – R) values (ppm) for the C-5 MTPA derivatives of cyclooctat-9-en-5,7-diol



^{13}C-NMR Δδ (S – R) values (ppm)



^{1}H-NMR Δδ (S – R) values (ppm)

Figure 66. ^{1}H NMR (500 MHz, $CDCl_3$) of (-)-casbene

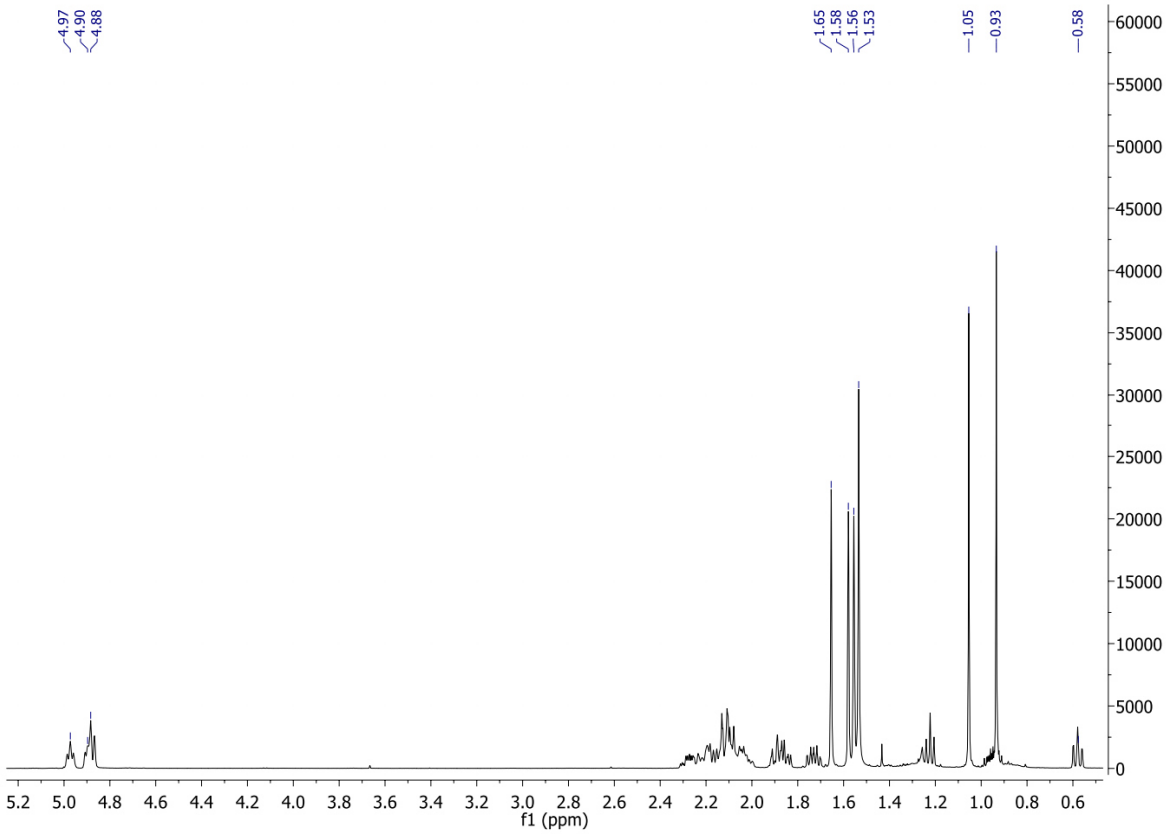


Figure 67. ^{13}C NMR (125 MHz, $CDCl_3$) of (-)-casbene

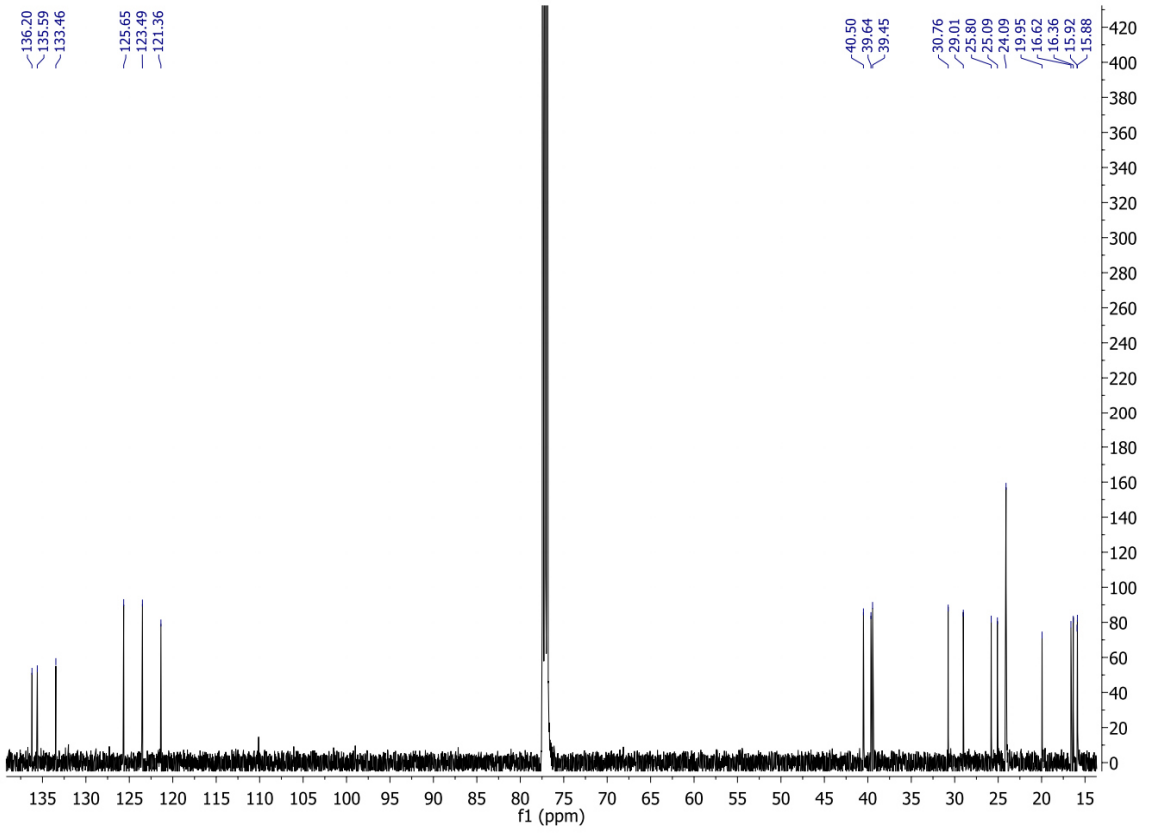


Figure 68. ^{1}H NMR ($CDCl_3$, 500 MHz, δ_H, J in Hz) of sinularcasbane D

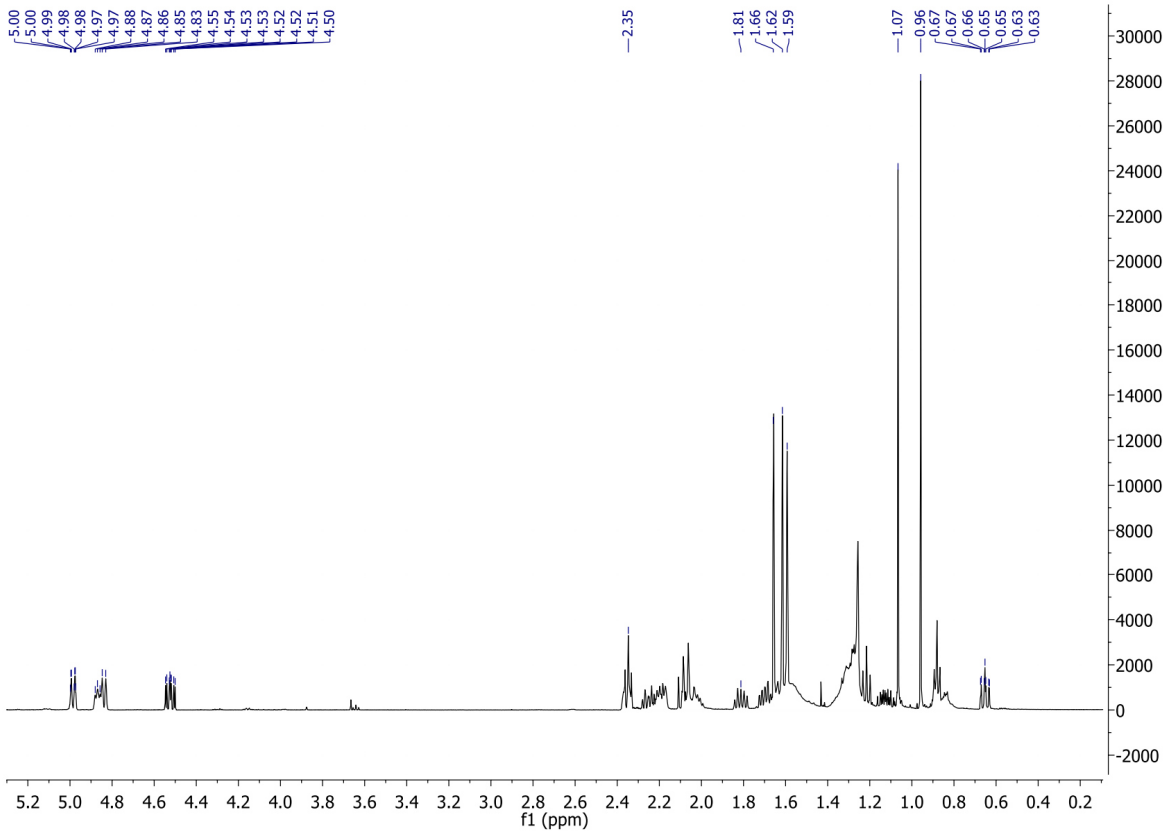


Figure 69. ^{1}H NMR (500 MHz, $CDCl_3$) of sinularcasbane D (olefinic protons)

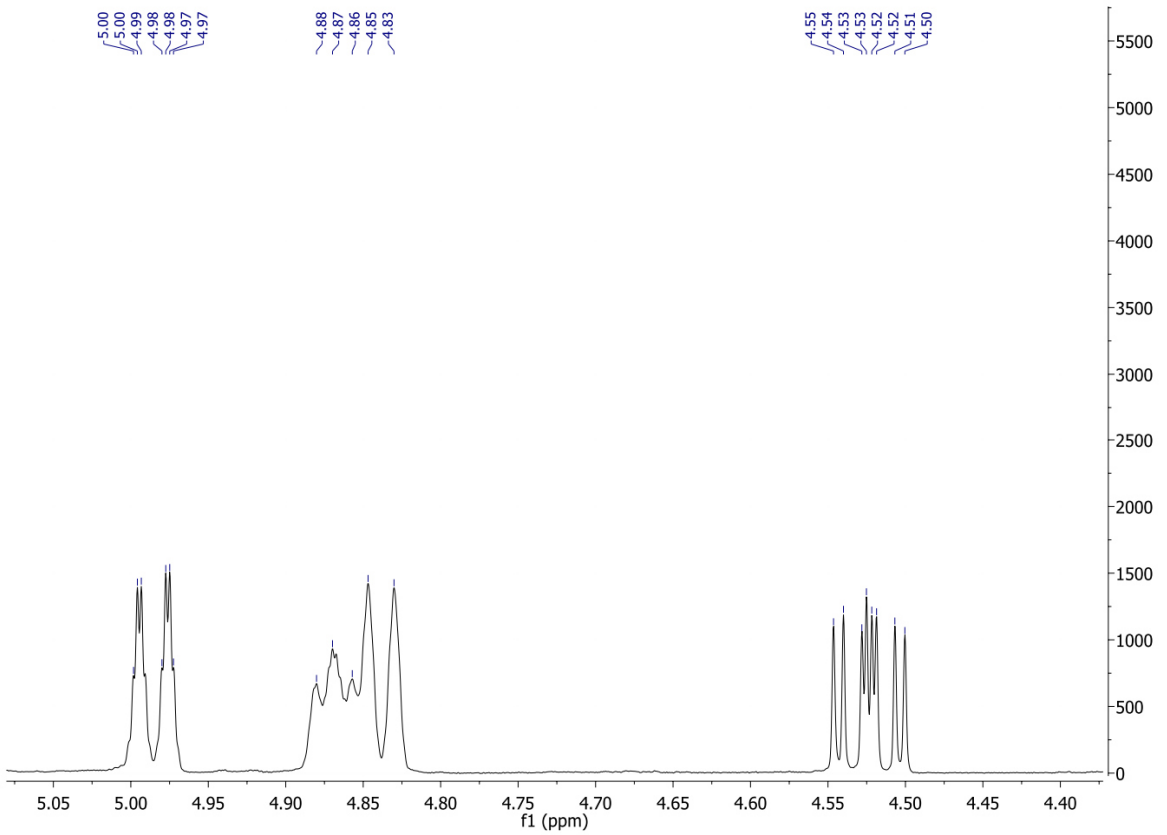


Figure 70. ^{13}C NMR (125 MHz, $CDCl_3$) of sinularcasbane D

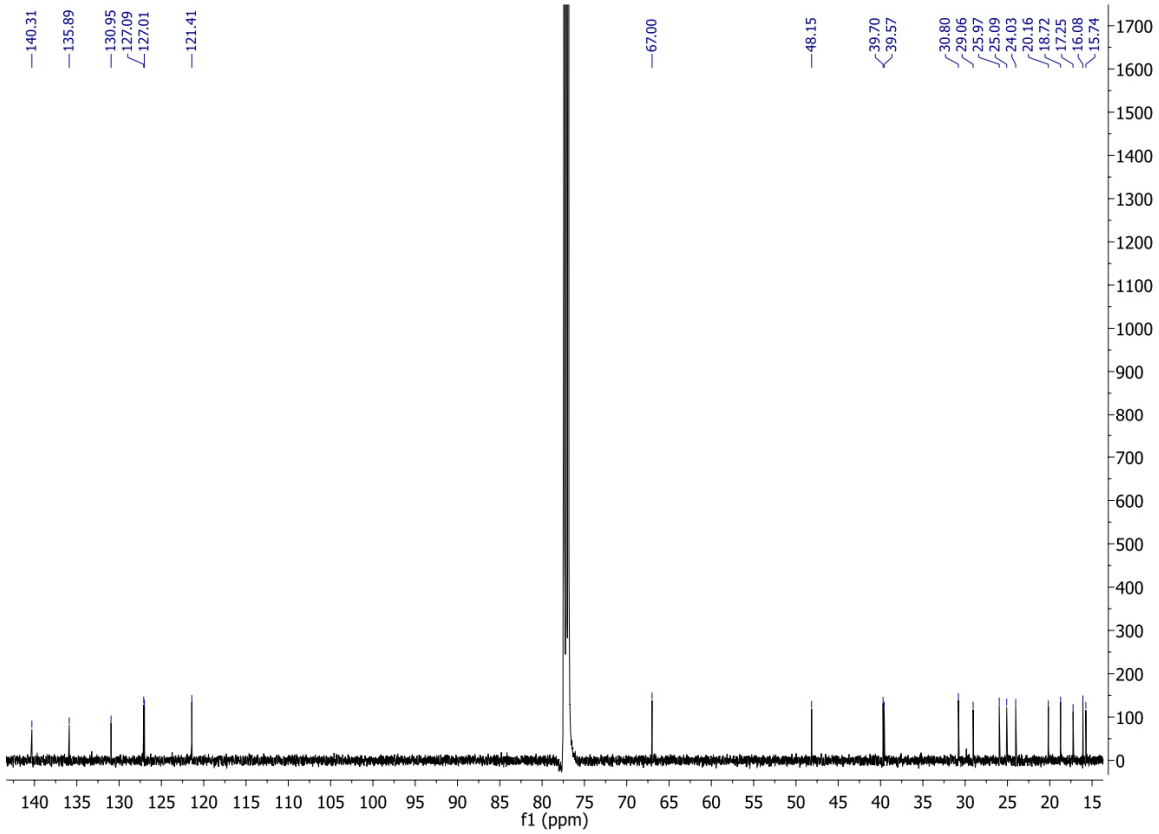


Figure 71. COSY NMR (500 MHz, $CDCl_3$) of sinularcasbane D

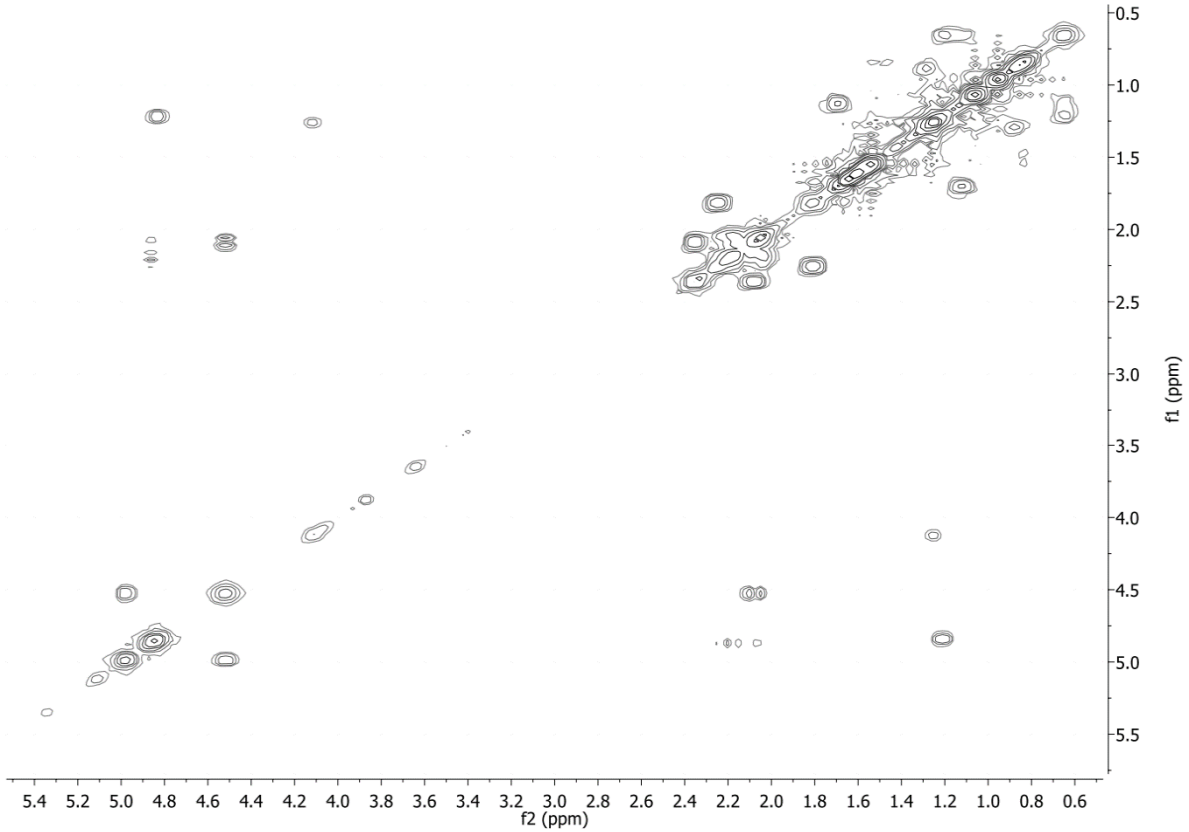


Figure 72 HSQC NMR (500 MHz, $CDCl_3$) of sinularcasbane D

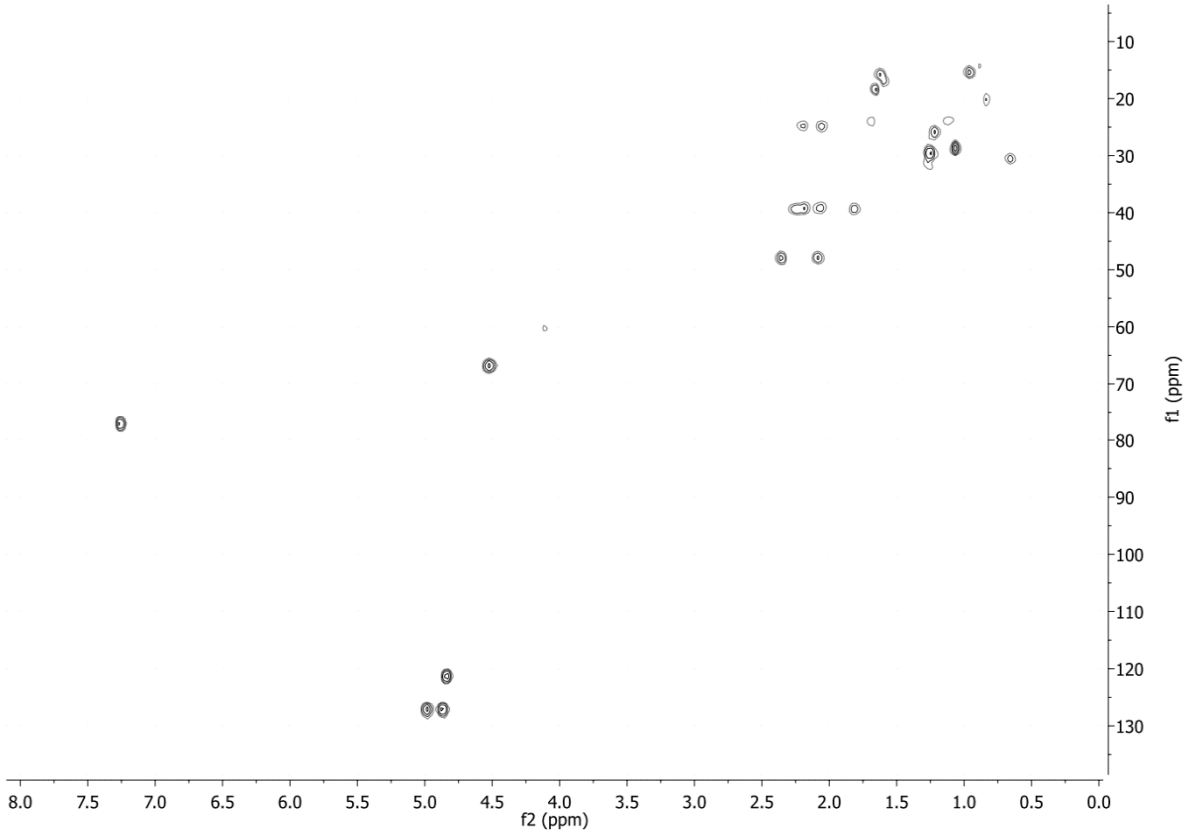


Figure 73. HMBC NMR (500 MHz, $CDCl_3$) of sinularcasbane D

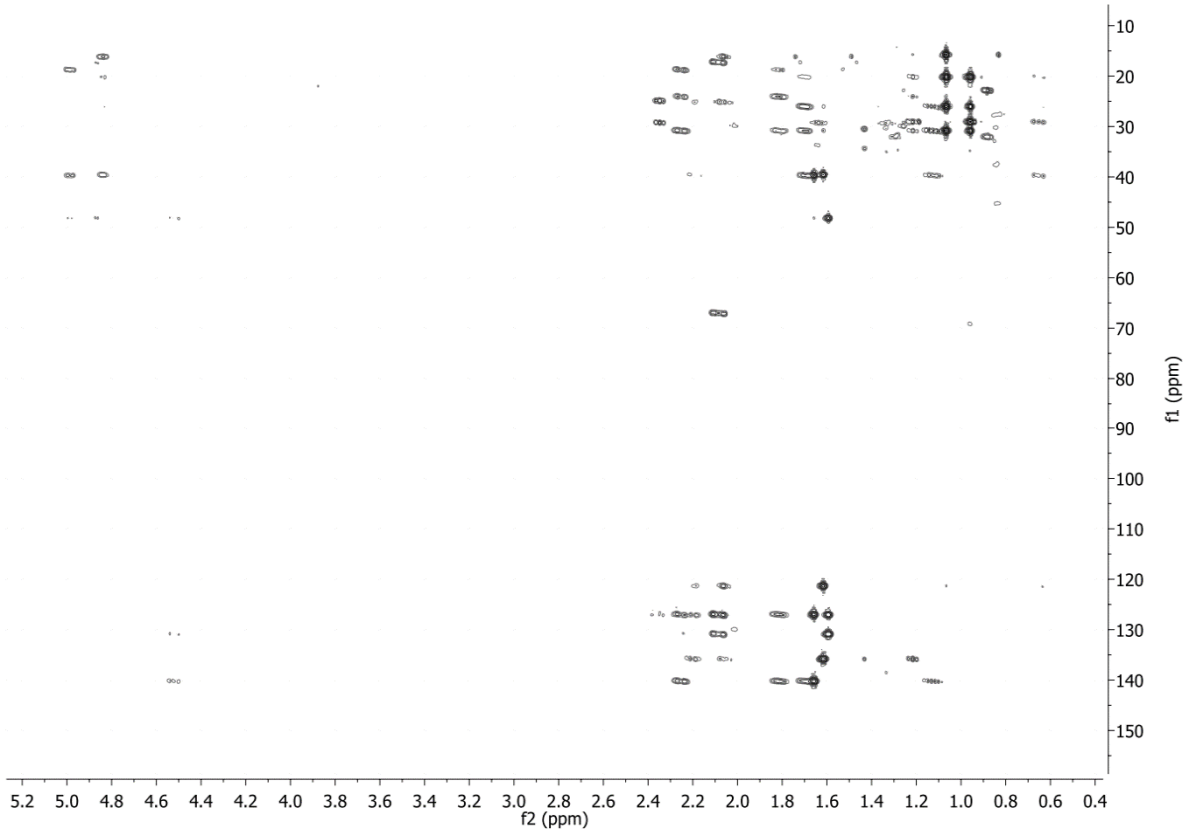


Figure 74 NOESY NMR (500 MHz, $CDCl_3$) of sinularcasbane D

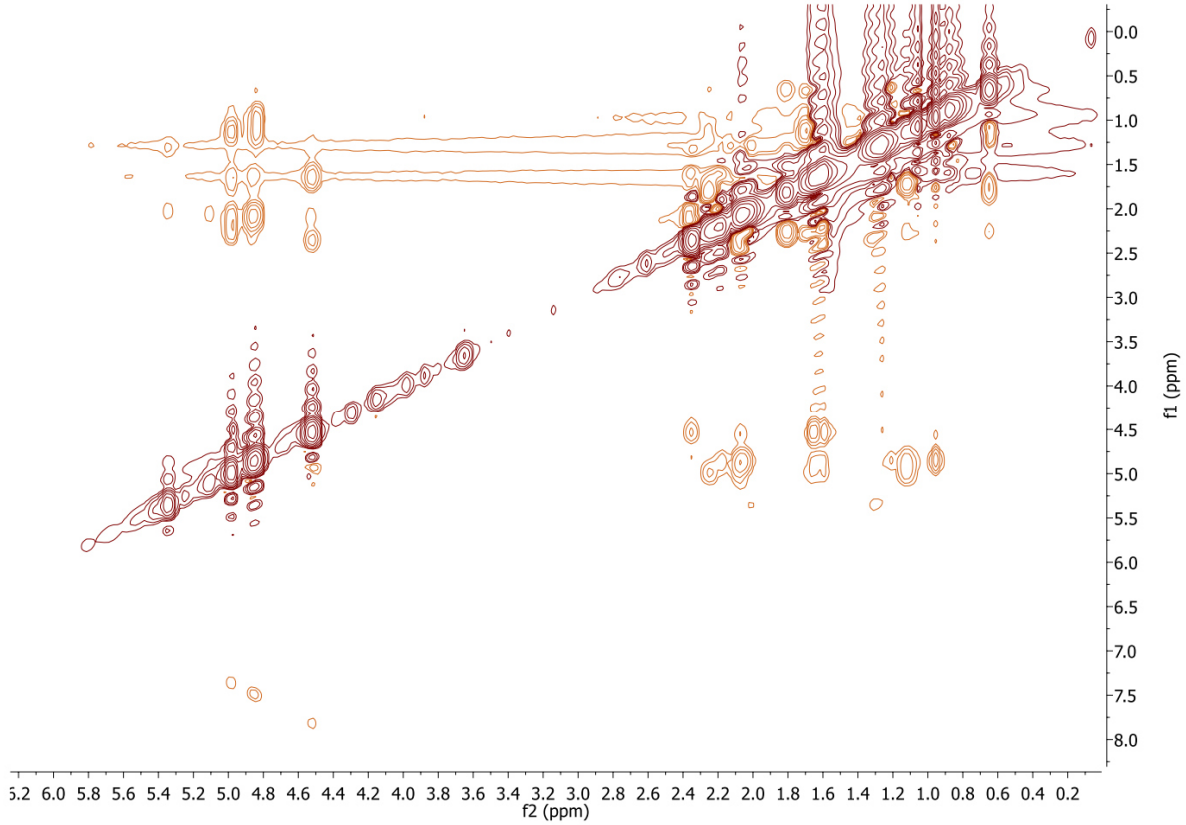


Figure 75. Circular dichroism spectroscopy of cyclooctat-9-en-7-ol (**a**), cyclooctat-9-en-5-7-diol (**b**), cyclooctatin (**c**)

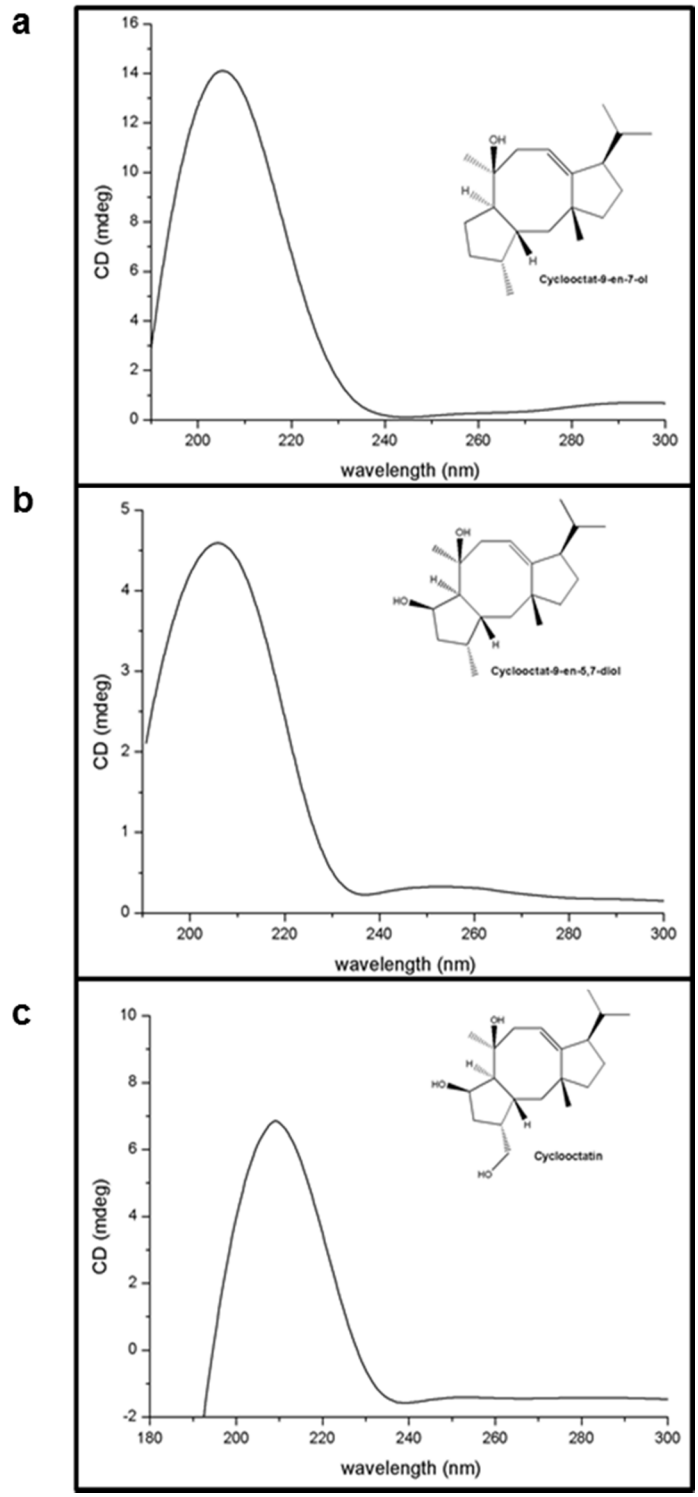


Figure 76. Circular dichroism spectroscopy of (1R,3E,7E,11S,12S)-3,7,18-dolabellatriene (**a**), R-cembrene A (**b**)

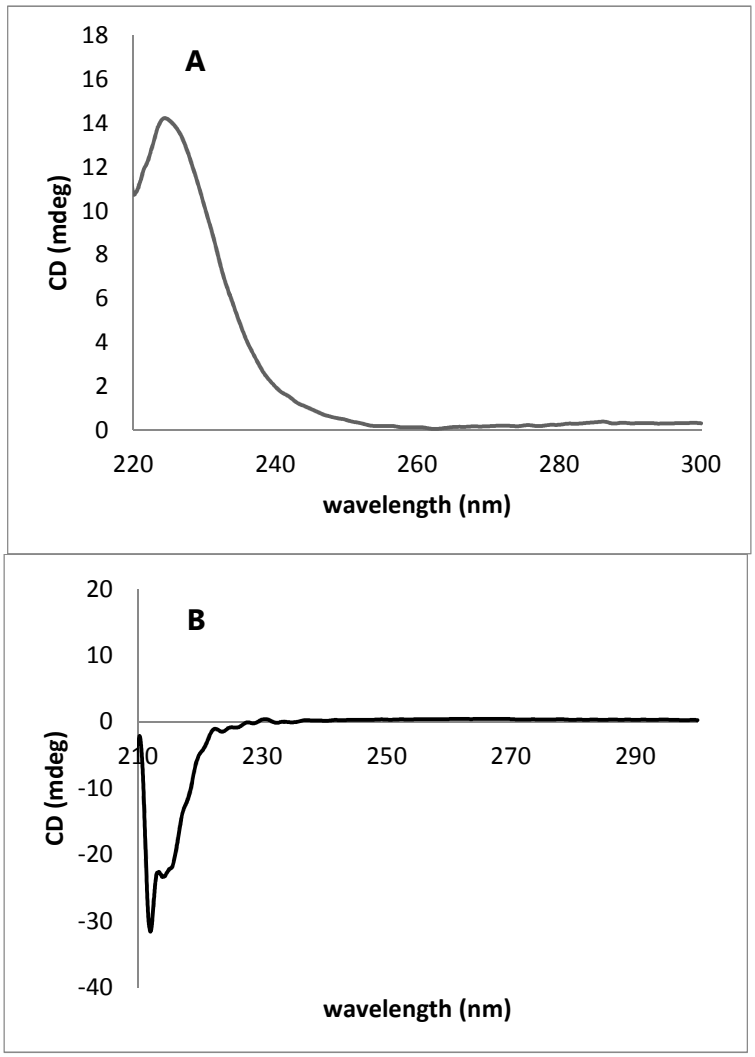


Figure 77. Mass spectrum of silylated cyclooctat-9-en-7-ol, recorded on a Trace GC Ultra with DSQII (Thermo Scientific), m/z was analyzed from [50-650]. m/z $C_{23}H_{42}OSi$ calculated. 362.30.

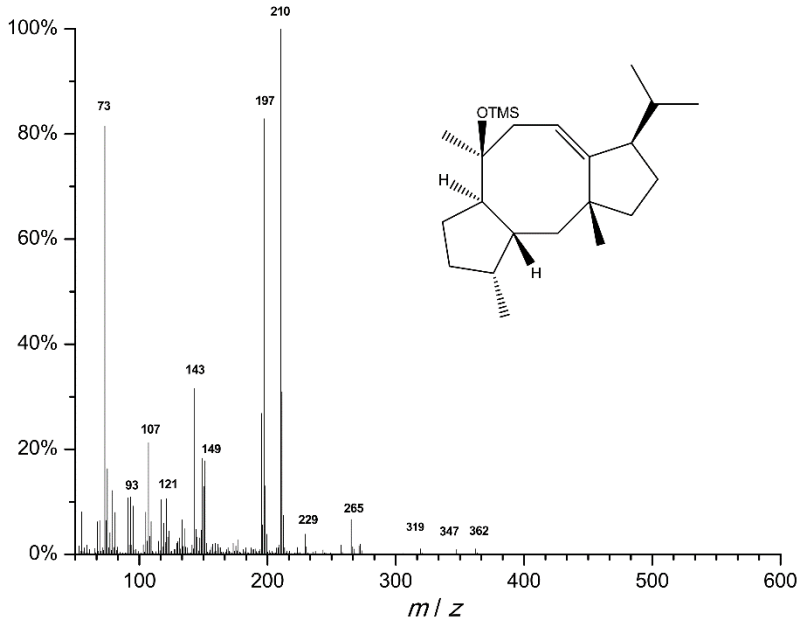


Figure 78. Mass spectrum of silylated cyclooctat-9-en-5,7-diol, recorded on a Trace GC Ultra with DSQII (Thermo Scientific), m/z was analyzed from [50-650]. *m/z* $C_{26}H_{50}O_2Si_2$ calculated 450.33.

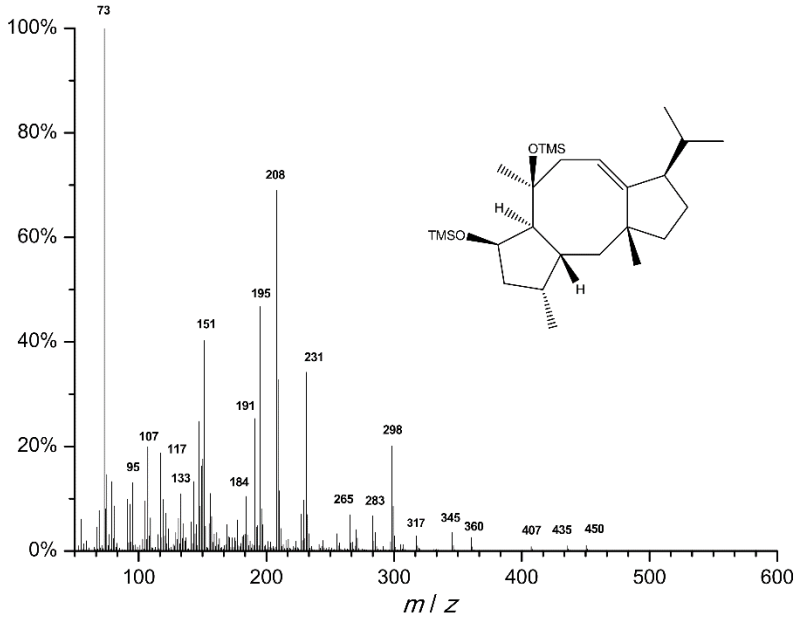


Figure 79. Mass spectrum of silylated cyclooctatin, recorded on a Trace GC Ultra with DSQII (Thermo Scientific), m/z was analyzed from [50-650]. *m/z* $C_{28}H_{56}O_3Si_3$ calculated 524.35.

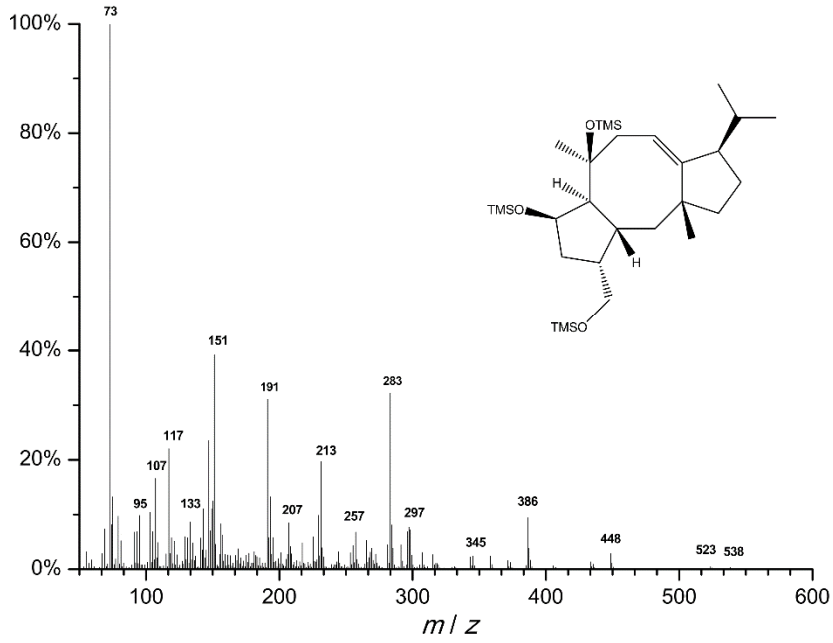


Figure 80 Mass spectrum of sinularcasbane D, recorded on a Trace GC Ultra with DSQII (Thermo Scientific), m/z was analyzed from [50-650]. *m/z* $C_{20}H_{32}O$ calculated 288.24.

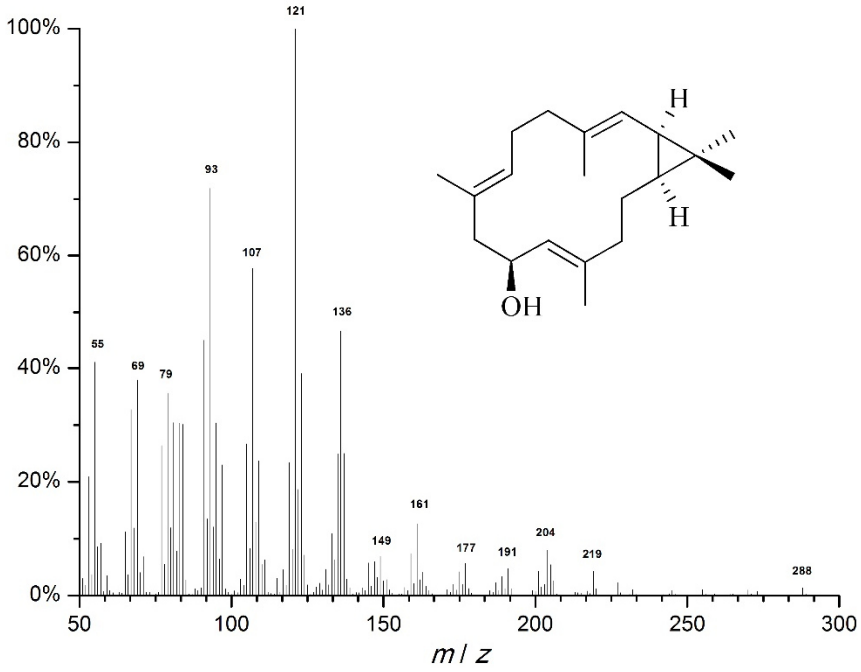


Figure 81. Gene of 1-deoxy-D-xylulose 5-phosphate synthase (*dxs*)

CCATGGATGAGTTTTGATATTGCCAAATACCCGACCCTGGCACTGGTCGACTCCACCCAGGAGTTACGACTGTTGCCGAAA
GAGAGTTTACCGAAACTCTGCGACGAACTGCGCCGCTATTTACTCGACAGCGTGAGCCGTTCCAGCGGGCACTTCGCCTCC
GGGCTGGGCACGGTCGAACTGACCGTGGCGCTGCACTATGTCTACAACACCCCGTTTGACCAATTGATTTGGGATGTGGG
GCATCAGGCTTATCCGCATAAAATTTTGACCGGACGCCGCGACAAAATCGGCACCATCCGTCAGAAAGGCGGCCTGCACC
CGTTCCCGTGGCGCGGCGAAAGCGAATATGACGTATTAAGCGTCGGGCATTCATCAACCTCCATCAGTGCCGGAATTGGT
ATTGCGGTTGCTGCCGAGAAAGAAGGCAAAAATCGCCGCACCGTCTGTGTCATTGGCGATGGCGCGATTACCGCTGGCAT
GGCGTTTGAAGCGATGAATCACGCGGGCGATATCCGTCCTGATATGCTGGTGGTCCTCAACGACAATGAAATGTCGATTTC
CGAAAATGTCGGCGCGCTCAATAACCATCTGGCACAGCTGCTTTCCGGTAAGCTTTACTCTTCGCTGCGCGAAGGCGGGAA
AAAAGTTTTCTCTGGCGTTCCGCCAATTAAAGAGCTGCTCAAACGTACCGAAGAACATATTAAAGGCATGGTAGTGCCTGG
CACGTTGTTTGAAGAGCTGGGCTTTAACTACATCGGCCCGGTTGACGGTCACGATGTGCTGGGGCTTATCACCACGCTGAA
GAACATGCGCGACCTGAAAGGCCCGCAGTTCCTGCATATCATGACCAAAAAAGGTCGTGGTTATGAACCGGCAGAAAAAG
ACCCCATCACTTTCCACGCCGTGCCTAAATTTGATCCCTCCAGCGGTTGTTTGCCGAAAAGTAGCGGCGGTTTGCCGAGCT
ATTCAAAAATCTTTGGCGACTGGTTGTGCGAAACGGCAGCGAAAGACAACAAGCTGATGGCGATTACTCCGGCGATGCGT
GAAGGTTCCGGCATGGTCGAGTTTTCACGTAAATTCCCGGATCGTTACTTCGACGTGGCAATCGCCGAGCAACACGCGGTG
ACCTTTGCCGCCGGTCTGGCGATTGGTGGGTACAAACCCATTGTCGCGATTTACTCCACTTTCCTGCAACGCGCCTATGATC
AGGTGCTGCATGACGTGGCGATTCAAAAGCTCCCGGTCCTGTTCGCCATCGACCGCGCGGGCATTGTTGGTGCTGACGGT
CAAACCCATCAGGGCGCTTTTGACCTCTCTTACCTGCGCTGTATACCGGAAATGGTCATTATGACCCCGAGCGATGAAAAC
GAATGTCGCCAGATGCTCTATACCGGCTATCACTATAACGACGGCCCGTCCGCGGTGCGCTACCCGCGCGGTAACGCGGT
TGGCGTGGAACTGACGCCGCTGGAAAAACTGCCAATTGGCAAAGGCATTGTGAAGCGTCGTGGCGAGAAACTGGCGATC
CTTAACTTTGGTACGCTGATGCCAGACGCGGCGAAAGTCGCTGAATCGCTGAACGCTACGCTGGTCGATATGCGTTTTGTG
AAACCGCTTGATGAAGCGTTAATTCTGGAAATGGCCGCCAGCCATGAAGCGCTGGTCACCGTAGAAGAAAACGCCATTAT
GGGCGGCGCAGGCAGCGGCGTGAACGAAGTGCTAATGGCCCATCGTAAACCAGTACCCGTGCTGAACATTGGCCTGCCT
GACTTCTTTATTCCACAAGGAACTCAGGAAGAAATGCGCGCCGAACTCGGCCTCGATGCCGCCGGTATGGAAGCCAAAAT
CAAGGCCTGGCTGGCATAAGAATTC

Figure 82. Gene of 1-deoxy-D-xylulose 5-phosphate reductoisomerase (*dxp*)

CATATGAAGCAACTCACCATTCTGGGCTCGACCGGCTCGATTGGTTGCAGCACGCTGGACGTGGTGCGCCATAATCCCGA
ACACTTCCGCGTAGTTGCGCTGGTGGCAGGCAAAAATGTCACTCGCATGGTAGAACAGTGCCTGGAATTCTCTCCCCGCTA
TGCCGTAATGGACGATGAAGCGAGTGCGAAACTTCTTAAAACGATGCTACAGCAACAGGGTAGCCGCACCGAAGTCTTAA
GTGGGCAACAAGCCGCTTGCGATATGGCAGCGCTTGAGGATGTTGATCAGGTGATGGCAGCCATTGTTGGCGCTGCTGGG
CTGTTACCTACGCTTGCTGCGATCCGCGCGGGTAAAACCATTTTGCTGGCCAATAAAGAATCACTGGTTACCTGCGGACGT
CTGTTTATGGACGCCGTAAAGCAGAGCAAAGCGCAATTGTTACCGGTCGATAGCGAACATAACGCCATTTTTCAGAGTTTA
CCGCAACCTATCCAGCATAATCTGGGATACGCTGACCTTGAGCAAAATGGCGTGGTGTCCATTTTACTTACCGGGTCTGGT
GGCCCTTTCCGTGAGACGCCATTGCGCGATTTGGCAACAATGACGCCGGATCAAGCCTGCCGTCATCCGAACTGGTCGAT
GGGGCGTAAAATTTCTGTCGATTCGGCTACCATGATGAACAAAGGTCTGGAATACATTGAAGCGCGTTGGCTGTTTAACGC
CAGCGCCAGCCAGATGGAAGTGCTGATTCACCCGCAGTCAGTGATTCACTCAATGGTGCGCTATCAGGACGGCAGTGTTC
TGGCGCAGCTGGGGGAACCGGATATGCGTACGCCAATTGCCCACACCATGGCATGGCCGAATCGCGTGAACTCTGGCGT
GAAGCCGCTCGATTTTTGCAAACTAAGTGCGTTGACATTTGCCGCACCGGATTATGATCGTTATCCATGCCTGAAACTGGC
GATGGAGGCGTTCGAACAAGGCCAGGCAGCGACGACAGCATTGAATGCCGCAAACGAAATCACCGTTGCTGCTTTTCTTG
CGCAACAAATCCGCTTTACGGATATCGCTGCGTTGAATTTATCCGTACTGGAAAAAATGGATATGCGCGAACCACAATGTG
TGGACGATGTGTTATCTGTTGATGCGAACGCGCGTGAAGTCGCCAGAAAAGAGGTGATGCGTCTCGCAAGCTGACTCGAG

Figure 83. Bi-cistronic operon of 2-C-methyl-D-erythriol 4-phosphate cytidyltransferase synthase (*ispD*) 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (*ispF*)

CCATGGATGGCAACCACTCATTTGGATGTTTGCGCCGTGGTTCCGGCGGCCGGATTTGGCCGTCGAATGCAAACGGAATGT
CCTAAGCAATATCTCTCAATCGGTAATCAAACCATTCTTGAACACTCGGTGCATGCGCTGCTGGCGCATCCCCGGGTGAAA
CGTGTCGTCATTGCCATAAGTCCTGGCGATAGCCGTTTTGCACAACTTCCTCTGGCGAATCATCCGCAAATCACCGTTGTAG
ATGGCGGTGATGAGCGTGCCGATTCCGTGCTGGCAGGTTTGAAAGCCGCTGGCGACGCGCAGTGGGTATTGGTGCATGAC
GCCGCTCGTCCTTGTCTGCATCAGGATGACCTCGCGCGATTGTTGGCGTTGAGCGAAACCAGCCGCACGGGAGGGATCCT
AGCCGCACCAGTGCGCGATACGATGAAACGTGCCGAACCGGGCAAAAATGCCATTGCTCATACCGTTGATCGCAACGGCT
TATGGCACGCGCTGACGCCGCAATTTTTCCCTCGTGAGCTGTTACATGACTGTCTGACGCGCGCTCTAAATGAAGGCGCGA
CTATTACCGACGAAGCCTCGGCGCTGGAATATTGCGGATTCCATCCTCAGTTGGTCGAAGGCCGTGCGGATAACATTAAAG
TCACGCGCCCGGAAGATTTGGCACTGGCCGAGTTTTACCTCACCCGAACCATCCATCAGGAGAATACATAAGCAGGAGCA
GGAGCAGAAGGAGGAGCAGGAATGCGAATTGGACACGGTTTTGACGTACATGCCTTTGGCGGTGAAGGCCCAATTATCAT
TGGTGGCGTACGCATTCCTTACGAAAAAGGATTGCTGGCGCATTCTGATGGCGACGTGGCGCTCCATGCGTTGACCGATGC
ATTGCTTGGCGCGGCGGCGCTGGGGGATATCGGCAAGCTGTTCCCGGATACCGATCCGGCATTTAAAGGTGCCGACAGCC
GCGAGCTGCTACGCGAAGCCTGGCGTCGTATTCAGGCGAAGGGTTATACCCTGGGCAACGTCGATGTCACTATCATCGCT
CAGGCACCGAAGATGTTGCCGCACATTCCACAAATGCGCGTATTTATTGCCGAAGATCTCGGCTGCCATATGGATGATGTT
AACGTGAAAGCCACTACTACGGAAAAACTTGGATTTACCGGACGTGGGGAAGGGATTGCCTGTGAAGCGGTGGCGCTACT
CATTAAGGCAACAAAATGAGAATTC

Figure 84. Gene of Isopentenyl-diphosphate delta isomerase (*Idi*)

CATATGCAAACGGAACACGTCATTTTATTGAATGCACAGGGAGTTCCCACGGGTACGCTGGAAAAGTATGCCGCACACAC
GGCAGACACCCGCTTACATCTCGCGTTCTCCAGTTGGCTGTTTAATGCCAAAGGACAATTATTAGTTACCCGCCGCGCACT
GAGCAAAAAAGCATGGCCTGGCGTGTGGACTAACTCGGTTTGTGGGCACCCACAACTGGGAGAAAGCAACGAAGACGCA
GTGATCCGCCGTTGCCGTTATGAGCTTGGCGTGGAAATTACGCCTCCTGAATCTATCTATCCTGACTTTCGCTACCGCGCCA
CCGATCCGAGTGGCATTGTGGAAAATGAAGTGTGTCCGGTATTTGCCGCACGCACCACTAGTGCGTTACAGATCAATGATG
ATGAAGTGATGGATTATCAATGGTGTGATTTAGCAGATGTATTACACGGTATTGATGCCACGCCGTGGGCGTTCAGTCCGT
GGATGGTGATGCAGGCGACAAATCGCGAAGCCAGAAAACGATTATCTGCATTTACCCAGCTTAAATAACTCGAG

Figure 85. Gene of *cotB2*

CATATGACCACCGGTCTGAGCACCGCAGGCGCACAGGATATTGGTCGTAGCAGCGTTCGTCCGTATCTGGAAGAATGTAC
CCGTCGTTTTCAAGAAATGTTTGATCGTCATGTTGTTACCCGTCCGACCAAAGTTGAACTGACCGATGCAGAACTGCGTGAA
GTTATTGATGATTGTAATGCAGCAGTTGCACCGCTGGGTAAAACCGTTAGTGATGAACGTTGGATTAGCTATGTTGGTGTTG
TTCTGTGGTCACAGAGTCCGCGTCATATTAAAGATATGGAAGCATTTAAAGCCGTGTGCGTTCTGAATTGTGTTACCTTTGTT
TGGGATGATATGGACCCTGCACTGCATGATTTTGGTCTGTTTCTGCCTCAGCTGCGTAAAATTTGCGAAAAATACTATGGTC
CGGAAGATGCCGAAGTTGCCTATGAAGCAGCACGTGCATTTGTTACCAGCGATCACATGTTTCGTGATAGCCCGATTAAAG
CAGCACTGTGTACCACCAGTCCGGAACAGTATTTTCGTTTTCGTGTTACCGATATTGGCGTGGATTTTTGGATGAAAATGAG
CTATCCGATTTATCGCCATCCGGAATTTACCGAACATGCAAAAACCAGCCTGGCAGCACGTATGACCACCCGTGGTCTGAC
CATTGTTAATGATTTCTATAGCTATGATCGCGAAGTTAGCCTGGGTCAGATTACCAATTGTTTTCGTCTGTGTGATGTGAGTG
ATGAAACCGCCTTTAAAGAATTTTTTCAGGCACGTCTGGATGACATGATCGAAGATATTGAATGCATCAAAGCGTTTGATCA
GCTGACACAGGATGTTTTTCTGGATCTGATTTATGGCAATTTTGTGTGGACCACCTCCAACAAACGTTATAAAACCGCAGTG
AATGATGTGAACAGCCGTATTCAATAACTCGAG

Figure 86. Bi-cistronic *cotB3/cotB4* operon

CATATGCGTGAACGTGGTCCGGTTACACCGGCAAAAAGCAGCGCACCGCCTGAACGTCCGTGGACCACCGGTACAGCAC
CGGGTAGCGTTCCGCTGCTGGGTCATACAATGGCACTGTGGCGTCGTCCGCTGCAATTTCTGGCAAGCCTGCCTGCACATG
GTGATCTGGTTGAAGTTCGTCTGGGTCCGAGCCGTGCATATCTGGCATGTCATCCGGAACTGGTTCGTCAGGTTCTGCTGA
ATCCGCGTGTTTTTGATAAAGGTGGCGTTTTCGATAAAGCACGTCAGCTGCTGGGCAATAGCCTGAGCGTTAGCCGTGGTG
AAGATCATCGTTATCAGCGTCGTATGATTCAGCCTGCATTTCATACCCCGAAAATTGCCGCATATACCGCAGCAGTTGCAG
ATGATACCCGTGCAGCAATTGGTAGCTGGGAACCGGGTCGTACCCTGGATATTAGCGATACCATGCATGCACTGCTGATGC
GTGTTGCAGCACGTACCCTGTTTAGCACCGGTATTGATGAAGCAACCATCGATGAAGCCCGTCATTGTCTGCGTATTGTTAG
TGATGGTATCTATAAACGTACGATGGCACCGCTGGGTATTATGGAAAAACTGCCGACACCGGGTAATCGTCGTTATGATCG
TGCAAATGCACGTCTGCGTCAGATTGTTGATGAAATGATTCGTGAACGTCGTCGTAGCGGTGCAGATCATGGCGATCTGCT
GAGCACCCTGCTGCGTGCAGAACATCCGGAAACCGGTAAAGGTCTGGATGATGGTGAAGTTCTGGATCAGGTTGTTACCT
TTCTGGTTGCAGGTAGCGAAACCACCGCAAGCACCCTGGCATTTGTTTTTCATCTGCTGGGTGCCCATCCGGAAGTTGAAA
AACGTGTTCATGCAGAAATCGATGAAATTCTGGAAGGTCGTAGCCCGACCTTTGAAGATCTGCCGAGCCTGGAATATACCC
GTGGTGTTATTACCGAAAGCCTGCGTCTGTATCCGCCTAGCTGGATGGCAATGCGTGTTACCGCAGCCGAAACCGAACTGG
GTGGTCGTACCGTTCCGGCAGGCACCATGATTCTGTATAGCGCACAGGCACTGCATCATAACCCGGAACTGTTTCCAGATC
CGGAACGTTTTGATCCTGAACGTTGGCTGGGTGATCGTGCCAAAGAAGTTGAACGCGGAGCACTGCTGCCGTTTGGTGCC
GGTAGCCATAAATGTATTGGTGATGTGCTGGCACTGACCGAAACAGCACTGATTGTTGCAACCATTGCAAGCCGTTGGCGT
CTGCGTCCGGTTCCGGGTACAACCCTGCGTCCTGAACCGAAAGCAACCCTGGAACCTGGTCCGCTGCCGATGGTTTGTGA
ACCGCGTTAACTAGTAGGAGGAAAACATCATGAAAGATTTTTTTCGTATGCGCACCGCACAGCAGCCTGCCACACGTCATT
GGCGTCATACCGTTGCACCGGGTGGTCTGCCGCTGGCAGGTCACGCTCTGCTGATGGCACGTAAACCTCTGCAATTCCTGG
CCTCACTGCCAGCCCACGGCGATCTGGTGGAACTGCGCCTGGGACCGCGTCCGGTGTATCTGCCGTGCCACCCTGAACTG
GTGCAGCAGGTACTGGTTAATGCACGTGTTTATGATACAGGCGGTCCGGTGAAAGAAAAAGCAAAACCGATTCTGGGTAA
TGGTCTGATTACCAGCGATTGGGCTGATCATCGTCGCCAGCGTCGCCTGGTTCAGCCAGCCTTTCACACCGCACGTATTGC
AAAGTATGCCGAAGTTATGGAACGTGAATGTGAAGCAGAAAGCACCGCATGGACCGCACGTCGTCCGATTGATGTTAGCC
ATGAAATGCTGGCCCTGACAGCACGCGTTACCGCACGTGCACTGTTTTCAACCGATATGGCTCCGCATGCCGTTGCAGAAA
TTCAGCATTGCCTGCCGATTGTTGTTGAAGGTGCATATCGTCAGGCAATTGATCCGACCGGTCTGCTGGCCAAACTGCCTCT
GGCAGCAAATCGTCGCTTTGATGATGCACTGGCACGTCTGAACCAGCTGATTGATCGCATGATTGATGATTACAAAGCCAG
TGATGATGGCGATCGTGGTGATGTTCTGAGCGCACTGTTTGCAGCACAGGATGATGAAACCGGTGGTACAATGAGCGATC
AAGAAATTCATGATCAGGTTATGACACTGCTGCTGGCTGGTATTGAAACCACAGCCAGCGCACTGACCTGGGCATGGTTTC
TGCTGGGACGTAATCCGGGTGCGGAAGCAGCACTGCATGCGGAAGTGGATGAAGTGCTGGGTGGCCGTGCACCGCGTTA
TGCAGATGTTCCGCGTCTGGCATATACACAGCGTGTGTTTAGCGAAGCCCTGCGCCTGTTTCCTCCGGCATGGCTGTTTACC
CGTACCACCACCGAAACCACGGAACTGGGAGGCCGTCGTCTGCCTCCGGCTTCAGATGTGCTGATTAGCCCGTATGTGCT
GCATCGTGATCCAGCACTGTTTCCGCGTCCTGATAGCTTTGACCCGGATCGTTGGCTGCCAGAACGCGCAAAAGAAGTAAC
ACGCGGTAGCTATCTGCCTTTTGGTGGTGGTTCACGTAAATGCATTGGCGACGTTTTTGGTATGACCGAAGCAACACTGGC
ACTGGCAGCGATTGCCGGTCGTTGGCGGATGCGTCCTATTCCTGGCACCAAAATTCGTCCGCGTCCGCAGATGAGCCTGA
CCGCAGGTCCTCTGCGCATGATTCCGGAACCTCGTTAACTCGAG

Figure 87. Bi-cistronic *afR/afx* operon

CCATGGGCCCTGCAAATGATGCCTGTGTTATTGTTGGTGCAGGTCTGGCAGGCGCAAAAGCAGCACAGGCACTGCGTGAA
GAGGGTTTTGATGGTCCGCTGGTTCTGATTGGTGATGAACGTGAACGTCCGTATGAACGTCCTCCGCTGAGCAAAGGTTAT
CTGACCGGTAAAGATGCACGTGAGCAGATTTATGTTCATCCGCCTCAGTGGTATGCAGAACATAATGTTGATATGCGTCTG
GGTATGGCAGTTACCGCAGTTGATCCGGCAGCACGTGAAATGACCCTGGATGATGGTAGCCGTGTTGGTTATGGTAAACT
GCTGCTGACCACCGGTAGCGCACCGCGTCGTCTGCCGGTTCCTGGTGCTGGCCTGGAACGTGTTCTGTATCTGCGTCGTGT
TGAAGATAGCGATCAGATTAAAGAAGCATTTCAGAGCGCAAGCCGTGCAGTTGTTATTGGTGCCGGTTGGATTGGTCTGGA
AACCACCGCAGCAGCACGTACCGCAGGCGTTGAAGTTACCGTTCTGGAAATGGCAGAACTGCCGCTGCTGCGTGTTCTGG
GTCGTGAAGTTGCACAGCTGTTTGCAAATCTGCATCGTGATCATGGTGTTGATCTGCGTTTTGGTGCACAGGTTGCAGAAAT
TACCGGTAGTGGTGGTGCAGTTGATGGTGTTCGTCTGAGTGATGGCACCCGTATTGATGCAGATGTTGTGATTGTGGGTGTT
GGCATTAAACCGAATATTGGCCTGGCACAAGAAGCCGGTCTGGAAGTTGATAATGGTATTCGTGTGGATGAACGTCTGCGT
ACCAGCTATCCTGATATTTATGCAGCCGGTGATGTTGCACATGCATTTCATCCGCTGCTGGGTAAACATATTCGCGTTGAAC

ATTGGGCAAATGCACTGAATCAGCCGCAGATTGCAGCAAAAGCAATGCTGGGTCGCGAAGATGCCGTTTATGATCGTATT
CCGTATTTTTTCACCGATCAGTATGATCTGGGCATGGAATATGCAGGTTATGTTGAACCGGGTGGTTATGATCAGGTTGTTT
TTCGTGGTGATGTGGCAGGTCGTGAATTTATTGCATTTTGGCTGGCAGGCAATCGTGTGCTGGCAGGTATGAATGTGAATAT
TTGGGATGTTAATGATCAGCTGCAGACCCTGGTTCGTACCGCACAGACCGTTGATATTCCGATGCTGACCGATCCGCAGGT
TCCGCTGGGTAGCCTGCTGCCTGATCCTCAGCATCGTTAATATATTAGTTAAGTATAAGAAGGAGATATAATGCCGAAAGTT
ACCTATGTTAGTGATGCCGGTGAAGTTCGTGTTGTTGATGGTCTGGTTGGTGATAGCGTTATGCAGACCGCAGTTCGTAATG
GTGTTCCGGGTATTACCGGTGAATGTGGTGGTGTTCTGAGCTGTGCAACCTGTCATGTTTTTGTTGATGAAGCAGATCTGGA
TCGTCTGGAACCGGTTAGCGGTCTGGAAGATGAAATGCTGGATGGCACCGTTGTTGATCGTTGTCCGAATAGCCGTCTGAG
CTGCCAGATTAAACTGAGCGAAGAACTGGGTGATCTGCGTGTTACCACACCGGAAGCACAAGAATAAGCGGCCGC

Figure 88. Bi-cistronic *pdR/pdx* operon

CCATGGGCAACGCCAATGATAATGTTGTTATTGTTGGCACCGGTCTGGCAGGCGTTGAAGTTGCATTTGGTCTGCGTGCAA
GCGGTTGGGAAGGTAATATTCGTCTGGTTGGTGATGCAACCGTTATTCCGCATCATCTGCCTCCGCTGAGCAAAGCATATCT
GGCAGGTAAAGCAACCGCAGAAAGCCTGTATCTGCGTACACCGGATGCCTATGCAGCACAGAATATTCAGCTGCTGGGTG
GTACACAGGTTACCGCAATTAATCGTGATCGTCAGCAGGTTATTCTGAGTGATGGTCGTGCACTGGATTATGATCGTCTGGT
GCTGGCAACCGGTGGTCGTCCGCGTCCGCTGCCGGTTGCAAGTGGTGCAGTTGGTAAAGCCAATAACTTTCGTTATCTGCG
CACCCTGGAAGATGCAGAATGTATTCGTCGTCAGCTGATTGCAGATAATCGCCTGGTTGTGATTGGTGGTGGTTATATTGGT
CTGGAAGTTGCAGCAACCGCCATTAAAGCAAATATGCATGTTACCCTGCTGGATACCGCAGCACGTGTTCTGGAACGTGTT
ACCGCACCGCCTGTTAGCGCCTTTTATGAACATCTGCATCGTGAAGCCGGTGTTGATATTCGTACCGGCACCCAGGTTTGTG
GTTTTGAAATGAGCACCGATCAGCAGAAAGTTACCGCAGTTCTGTGTGAAGATGGCACCCGTCTGCCTGCAGATCTGGTTA
TTGCAGGTATTGGCCTGATTCCGAATTGTGAACTGGCAAGCGCAGCAGGTCTGCAGGTTGATAATGGTATTGTTATTAACG
AACACATGCAGACCAGCGATCCGCTGATTATGGCAGTTGGTGATTGTGCACGTTTTCATAGCCAGCTGTATGATCGTTGGG
TTCGTATTGAAAGCGTTCCGAATGCACTGGAACAGGCACGTAAAATTGCAGCAATTCTGTGTGGTAAAGTTCCGCGTGATG
AAGCAGCACCGTGGTTTTGGAGCGATCAGTATGAAATTGGTCTGAAAATGGTTGGTCTGAGCGAAGGTTATGATCGCATTA
TTGTTCGTGGTAGCCTGGCACAGCCGGATTTTTCAGTTTTTTATCTGCAGGGTGATCGTGTGCTGGCAGTTGATACCGTTAAT
CGTCCGGTTGAATTTAACCAGAGCAAACAAATTATCACCGATCGTCTGCCGGTGGAACCGAATCTGCTGGGAGATGAAAG
CGTGCCGCTGAAAGAAATTATTGCAGCAGCAAAAGCAGAACTGAGCAGCGCATAATATATTAGTTAAGTATAAGAAGGAG
ATATAATGAGCAAAGTTGTGTATGTTAGCCATGATGGCACCCGTCGTGAACTGGATGTTGCAGATGGTGTGAGCCTGATGC
AGGCAGCAGTTAGCAATGGTATTTATGATATTGTTGGTGATTGTGGTGGTAGCGCAAGCTGTGCAACCTGTCATGTTTATGT
TAATGAAGCCTTCACCGATAAAGTTCCGGCAGCAAATGAACGTGAAATTGGTATGCTGGAATGTGTTACCGCAGAACTGAA
ACCGAATAGCCGTCTGTGTTGTCAGATTATTATGACACCGGAACTGGACGGTATTGTTGTTGATGTTCCGGATCGTCAGTGG
TAAGCGGCCGC

Figure 89. (-)-Casbene synthase (*cs*) from *Jatropha curcas*

CATATGGCAAGCACCAAAAGCGAAACCGAAGCACGTCCGCTGGCATATTTTCCGCCTACCGTTTGGGGTGATCGTCTGGC
AAGCCTGACCTTTAATCAGCCTGCATTTGAACTGCTGAGTAAACAGGTTGAGCTGCTGAACGAGAAAATCAAAAAAGAAAT
GCTGAATGTGAGCACCAGCGATCTGGCAGAAAAAATCATTCTGATTGATAGCCTGTGTCGTCTGGGTGTTAGCTATCATTTT
GAAGAAGAAATCCAAGAAAACCTGACCCGCATCTTTAATACCCAGCCGAATTTTCTGAACGAAAAAGATTATGATTTATTC
ACCGTGGCCGTGATCTTTCGTGTTTTTCGTCAGCATGGCTTTAAAATCAGCAGCGACGTTTTTAACAAATTCAAAGATAGCG
ACGGGAAATTCAAAGAAAGCCTGCTGAATGATATTAAAGGCATCCTGAGCCTGTTTGAAGCAACCCATGTTAGCATGCCG
AATGAACCGATTCTGGATGAAGCACTGGCATTTACCAAAGCATTTCTGGAAAGCAGCGCAGTTAAATCCTTTCCGAATTTTG
CCAAACATATTAGCAGCGCACTGGAACAGCCGGTTCATAAAGGTATTCCGCGTCTGGAAGCACGCAAATATATCGATCTGT
ATGAAGTTGATGAAAGCCGCAATGAAACCGTGCTGGAACTGGCGAAACTGGATTTTAATCGTGTTCAGCTGCTGCATCAAG
AAGAACTGAGCCAGTTTAGCAAATGGTGGAAAAGCCTGAATATTAGTGCCGAAGTTCCGTATGCACGTAATCGTATGGCA
GAAATCTTTTTTTGGGCAGTGAGCATGTATTTTGAACCGCAGTATGCAAAAGCCCGTATGATTGTTAGCAAAGTGGTTCTGC
TGATTAGCCTGATTGATGATACCATTGATGCCTATGCCACCATCGATGAAATCCATCGTGTTGCAGATGCAATTGAACGTTG
GGATATGCGTCTGGTTGATCAGCTGCCGAATTATATGAAAGTGATTTATCGCCTGATCATCAACACCTTTGATGAGTTTGAA
AAAGATCTGGAAGCCGAGGGTAAAAGCTATAGCGTTAAATATGGTCGTGAAGCCTATCAAGAACTGGTGCGTGGTTATTA

CCTGGAAGCAATTTGGAAAGCGGATGGTAAAGTTCCGAGCTTCGATGAGTATATCTATAATGGTGGTGTTACCACCGGTCT
GCCGCTGGTTGCCACCGTTAGCTTTATGGGTGTTAAAGAAATCAAAGGCACCAAAGCCTTTCAGTGGCTGAAAACCTATCC
GAAACTGAATCAGGCAGGCGGTGAATTTATTCGTCTGGTTAATGATGTTATGAGCCATGAAACCGAACAGGATCGTGGTCA
TGTTGCAAGCTGTATTGATTGCTATATGAAACAGTATGGCGTGAGCAAAGAAGAAGCCGTTGAAGAGATCCAGAAAATGG
CAACCAATGAGTGGAAAAAACTGAACGAACAGCTGATTGTTCGTAGCACCGAAGTTGTTCCGGTTAATCTGCTGATGCGTA
TTGTTAATCTGGTTCGTCTGACCGATGTGAGCTATAAATACGGTGATGGTTATACCGATAGCTCCCAGCTGAAAGAATATGT
GAAAGGCCTGTTTATTGAACCGATCGCAACCGGTACCTCTTAACTCGAG

Figure 90. Taxadiene synthase (*txs*) from *Taxus brevifolia*

CATATGGCAATGAGCAGCAGCACCGGCACCAGCAAAGTTGTTAGCGAAACCAGCAGTACCATTGTTGATGATATTCCGCG
TCTGAGCGCAAATTATCATGGTGATCTGTGGCATCATAATGTGATTCAGACCCTGGAAACCCCGTTTCGTGAAAGCAGCAC
CTATCAAGAACGTGCAGATGAACTGGTTGTGAAAATCAAAGATATGTTTAACGCACTGGGTGATGGTGATATTAGCCCGAG
CGCCTATGATACCGCATGGGTTGCACGTCTGGCAACCATTAGCAGTGATGGTAGCGAAAAACCGCGTTTTCCGCAGGCACT
GAATTGGGTTTTTAACAATCAGCTGCAGGATGGTAGTTGGGGTATTGAAAGCCATTTTAGCCTGTGTGATCGTCTGCTGAAT
ACCACCAATAGCGTTATTGCACTGAGCGTTTGGAAAACCGGTCATAGCCAGGTTCAGCAGGGTGCAGAATTTATTGCAGAA
AATCTGCGCCTGCTGAATGAAGAAGATGAGCTGAGTCCGGATTTTCAGATTATCTTTCCGGCACTGCTGCAGAAAGCAAAA
GCACTGGGTATTAATCTGCCGTATGATCTGCCGTTTATCAAATATCTGAGCACCACCCGTGAAGCACGTCTGACCGATGTTA
GCGCAGCAGCAGATAATATTCCGGCAAATATGCTGAATGCACTGGAAGGTCTGGAAGAAGTTATTGACTGGAACAAAATT
ATGCGCTTCCAGAGCAAAGATGGTAGCTTTCTGAGTAGTCCGGCAAGCACCGCATGTGTTCTGATGAATACCGGTGATGAA
AAATGCTTTACCTTCCTGAATAACCTGCTGGATAAATTTGGTGGTTGTGTTCCGTGTATGTATAGCATTGATCTGCTGGAACG
TCTGAGCCTGGTTGATAATATTGAACATCTGGGTATTGGTCGCCACTTCAAACAAGAAATTAAAGGTGCACTGGATTACGT
GTATCGTCATTGGAGCGAACGTGGTATTGGTTGGGGTCGTGATAGCCTGGTTCCGGATCTGAATACAACCGCACTGGGCCT
GCGTACCCTGCGTATGCATGGTTATAATGTTAGCTCAGATGTGCTGAACAACTTTAAAGATGAAAACGGTCGCTTTTTTAGC
AGCGCAGGTCAGACCCATGTTGAACTGCGTAGCGTTGTTAACCTGTTTCGTGCAAGCGATCTGGCATTTCCGGATGAACGT
GCAATGGATGATGCACGTAAATTTGCAGAACCGTATCTGCGTGAAGCCCTGGCCACCAAAATTAGCACCAATACAAAACT
GTTTAAAGAAATCGAATATGTGGTCGAGTATCCGTGGCACATGAGCATTCCTCGTCTGGAAGCACGTAGCTATATTGATAG
CTATGATGATAACTATGTGTGGCAGCGTAAAACCCTGTATCGTATGCCGAGCCTGAGCAATAGCAAATGTCTGGAACTGGC
AAAACTGGATTTTAACATTGTTCAGAGCCTGCACCAAGAAGAACTGAAACTGCTGACCCGTTGGTGGAAAGAAAGCGGTA
TGGCAGATATTAACTTTACCCGTCATCGTGTTGCCGAAGTGTATTTTAGCAGTGCAACCTTTGAACCGGAATATAGCGCAAC
CCGTATTGCCTTTACCAAAATTGGTTGTCTGCAGGTCCTGTTCGATGATATGGCCGATATTTTTGCAACCCTGGATGAACTG
AAAAGTTTTACCGAAGGTGTTAAACGTTGGGATACCAGTCTGCTGCATGAAATCCCGGAATGTATGCAGACCTGTTTTAAA
GTGTGGTTTAAACTGATGGAAGAGGTGAATAACGATGTGGTTAAAGTTCAGGGTCGCGATATGCTGGCCCATATTCGTAAA
CCGTGGGAACTGTATTTCAACTGCTATGTTCAAGAACGCGAATGGCTGGAAGCCGGTTATATTCCGACCTTTGAAGAATAT
CTGAAAACCTATGCAATTAGCGTTGGTCTGGGTCCGTGTACCCTGCAGCCGATTCTGCTGATGGGTGAACTGGTGAAAGAT
GATGTTGTTGAGAAAGTTCATTACCCGAGCAACATGTTTGAACTGGTAAGCCTGAGCTGGCGTCTGACCAATGATACCAAA
ACCTATCAGGCAGAAAAAGCACGTGGTCAGCAGGCAAGCGGTATTGCATGTTATATGAAAGACAATCCGGGTGCAACCGA
AGAGGATGCAATCAAACATATTTGTCGTGTTGTTGATCGTGCACTGAAAGAAGCCAGCTTTGAATATTTCAAACCGAGCAA
CGATATTCCGATGGGCTGTAAATCCTTTATCTTTAATCTGCGTCTGTGCGTGCAGATCTTCTATAAATTCATTGATGGTTACG
GCATTGCCAACGAAGAGATCAAAGATTATATCCGCAAAGTGTATATCGATCCGATTCAGGTTTAACTCGAG

9. Literature

1 J. M. Ageitos, J. A. Vallejo, P. Veiga-Crespo and T. G. Villa, *Appl. Microbiol. Biotechnol.*, 2011, **90**, 1219–27.
2 P. Gujjari, S. O. Suh, K. Coumes and J. J. Zhou, *Mycologia*, 2011, **103**, 1110–1118.
3 S. Wu, C. Hu, X. Zhao and Z. K. Zhao, *Eur. J. Lipid Sci. Technol.*, 2010, **112**, 727–733.
4 T. C. Adarme-Vega, S. R. Thomas-Hall and P. M. Schenk, *Curr. Opin. Biotechnol.*, 2014, **26**, 14–8.
5 Z. Xue, P. L. Sharpe, S.-P. Hong, N. S. Yadav, D. Xie, D. R. Short, H. G. Damude, R. a Rupert, J. E. Seip, J. Wang, D. W. Pollak, M. W. Bostick, M. D. Bosak, D. J. Macool, D. H. Hollerbach, H. Zhang, D. M. Arcilla, S. a Bledsoe, K. Croker, E. F. McCord, B. D. Tyreus, E. N. Jackson and Q. Zhu, *Nat. Biotechnol.*, 2013, **31**, 734–40.
6 M. Adamczak, U. T. Bornscheuer and W. Bednarski, *Eur. J. Lipid Sci. Technol.*, 2008, **110**, 491–504.
7 D. W. Christianson, *Chem. Rev.*, 2006, **106**, 3412–42.
8 T. J. Maimone and P. S. Baran, *Nat. Chem. Biol.*, 2007, **3**, 396–407.
9 T. Brück, R. Kourist and B. Loll, *ChemCatChem*, 2014, **6**, 1142–1165.
10 S.-Y. Kim, P. Zhao, M. Igarashi, R. Sawa, T. Tomita, M. Nishiyama and T. Kuzuyama, *Chem. Biol.*, 2009, **16**, 736–43.
11 E. Ioannou, A. Quesada, M. M. Rahman, S. Gibbons, C. Vagias and V. Roussis, *J. Nat. Prod.*, 2011, **74**, 213–22.
12 I. F. Sevrioukova and T. L. Poulos, *Arch. Biochem. Biophys.*, 2011, **507**, 66–74.
13 T. Aoyama, H. Naganawa, Y. Muraoka, T. Aoyagi and T. Takeuchi, *J. Antibiot. (Tokyo).*, 1992, **45**, 1703–4.
14 J. Yin, M. Zhao, M. Ma, Y. Xu, Z. Xiang, Y. Cai, J. Dong, X. Lei, K. Huang and P. Yan, *Mar. Drugs*, 2013, **11**, 455–65.
15 P. Smith, *Glob. Food Sec.*, 2013, **2**, 18–23.
16 S. Chu and A. Majumdar, *Nature*, 2012, **488**, 294–303.
17 A. E. Ercin and A. Y. Hoekstra, *Environ. Int.*, 2014, **64**, 71–82.
18 C. D. Thomas, A. Cameron, R. E. Green, M. Bakkenes, L. J. Beaumont, Y. C. Collingham, B. F. N. Erasmus, M. F. De Siqueira, A. Grainger, L. Hannah, L. Hughes, B. Huntley, A. S. Van Jaarsveld, G. F. Midgley, L. Miles, M. A. Ortega-Huerta, A. T. Peterson, O. L. Phillips and S. E. Williams, *Nature*, 2004, **427**, 145–8.
19 A. R. Morais and R. Bogel-Lukasik, *Sustain. Chem. Process.*, 2013, **1**, 18.
20 R. Hatti-Kaul, U. Törnvall, L. Gustafsson and P. Börjesson, *Trends Biotechnol.*, 2007, **25**, 119–24.
21 J. H. Clark, V. Budarin, F. E. I. Deswarte, J. J. E. Hardy, F. M. Kerton, A. J. Hunt, R. Luque, D. J. Macquarrie, K. Milkowski, A. Rodriguez, O. Samuel, S. J. Tavener, R. J. White and A. J. Wilson, *Green Chem.*, 2006, **8**, 853.
22 J. H. Clark, F. E. I. Deswarte and T. J. Farmer, *Biofuels, Bioprod. Biorefining*, 2009, **3**, 72–90.
23 M. M. Wright and R. C. Brown, *Biofuels, Bioprod. Biorefining*, 2007, **1**, 49–56.
24 S. Wenda, S. Illner, A. Mell and U. Kragl, *Green Chem.*, 2011, 3007–3047.
25 N. Ran, L. Zhao, Z. Chen and J. Tao, *Green Chem.*, 2008, **10**, 361.
26 J. H. Clark, *Green Chem.*, 1999, **1**, 1–8.
27 P. T. Anastas and J. B. Zimmerman, *Environ. Sci. Technol.*, 2003, **37**, 94A–101A.
28 J. H. Clark, *Green Chem.*, 2006, **8**, 17–21.
29 R. H. Wijffels, M. J. Barbosa and M. H. M. Eppink, *Biofuels, Bioprod. Biorefining*, 2010, **4**, 287–295.
30 S. K. Lee, H. Chou, T. S. Ham, T. S. Lee and J. D. Keasling, *Curr. Opin. Biotechnol.*, 2008, **19**, 556–63.
31 N. Ferrer-Miralles, J. Domingo-Espín, J. L. Corchero, E. Vázquez and A. Villaverde, *Microb. Cell Fact.*, 2009, **8**, 17.
32 M. C. Y. Chang and J. D. Keasling, *Nat. Chem. Biol.*, 2006, **2**, 674–81.
33 E. J. Steen, Y. Kang, G. Bokinsky, Z. Hu, A. Schirmer, A. McClure, S. B. Del Cardayre and J. D. Keasling, *Nature*, 2010, **463**, 559–62.
34 Y. Waché, M. Aguedo, J.-M. Nicaud and J.-M. Belin, *Appl. Microbiol. Biotechnol.*, 2003, **61**, 393–404.
35 J. H. Martens, H. Barg, M. J. Warren and D. Jahn, *Appl. Microbiol. Biotechnol.*, 2002, **58**, 275–85.
36 F. Cherubini, *Energy Convers. Manag.*, 2010, **51**, 1412–1421.
37 A. Azapagic, *Trends Biotechnol.*, 2014, **32**, 1–4.
38 Grant Agreement 289284, in *www.chibiofp7.eu*.
39 P. D. V. S. Dr. Lars O. Wiemann, *GitLaborportal*, 2013.
40 M. N. . Ravi Kumar, *React. Funct. Polym.*, 2000, **46**, 1–27.
41 S. Papanikolaou and G. Aggelis, *Eur. J. Lipid Sci. Technol.*, 2011, **113**, 1031–1051.
42 J. P. Wynn and C. Ratledge, *Bailey's Industrial Oil and Fat Products*, John Wiley & Sons, Inc., Hoboken, NJ, USA, 2005.
43 C. Ratledge, *Biochimie*, 2004, **86**, 807–15.
44 H. M. Alvarez and a Steinbüchel, *Appl. Microbiol. Biotechnol.*, 2002, **60**, 367–76.
45 Q. Li, W. Du and D. Liu, *Appl. Microbiol. Biotechnol.*, 2008, **80**, 749–56.

46 T. Czabany, K. Athenstaedt and G. Daum, *Biochim. Biophys. Acta*, 2007, **1771**, 299–309.
47 C. Huang, X. Chen, L. Xiong, X. Chen, L. Ma and Y. Chen, *Biotechnol. Adv.*, **31**, 129–39.
48 Y. Chisti, *Trends Biotechnol.*, 2008, **26**, 126–31.
49 M. Hassan, P. Blanc, L. Granger and A. Pareilleux, *Process*, 1996, **31**, 355–361.
50 S. Wu, X. Zhao, H. Shen, Q. Wang and Z. K. Zhao, *Bioresour. Technol.*, 2011, **102**, 1803–7.
51 S. Wu, C. Hu, G. Jin, X. Zhao and Z. K. Zhao, *Bioresour. Technol.*, 2010, **101**, 6124–9.
52 C. Benning, Z. H. Huang and D. A. Gage, *Arch. Biochem. Biophys.*, 1995, **317**, 103–11.
53 P. J. le B. Williams, *Nature*, 2007, **450**, 478.
54 L. Gouveia and A. C. Oliveira, *J. Ind. Microbiol. Biotechnol.*, 2009, **36**, 269–74.
55 I. Lang, L. Hodac, T. Friedl and I. Feussner, *BMC Plant Biol.*, 2011, **11**, 124.
56 T. E. Lewis, P. D. Nichols and T. A. McMeekin, *Mar. Biotechnol.*, 1999, **1**, 580–587.
57 L. Rodolfi, G. Chini Zittelli, N. Bassi, G. Padovani, N. Biondi, G. Bonini and M. R. Tredici, *Biotechnol. Bioeng.*, 2009, **102**, 100–12.
58 G. Breuer, P. P. Lamers, D. E. Martens, R. B. Draaisma and R. H. Wijffels, *Bioresour. Technol.*, 2012, **124**, 217–26.
59 N. L. Hockin, T. Mock, F. Mulholland, S. Kopriva and G. Malin, *Plant Physiol.*, 2012, **158**, 299–312.
60 T. Mock and B. M. A. Kroon, *Phytochemistry*, 2002, **61**, 41–51.
61 N. Zhukova, *Phytochemistry*, 1995, **39**, 351–356.
62 W. R. Barclay, K. M. Meager and J. R. Abril, *J. Appl. Phycol.*, 1994, **6**, 123–129.
63 J. L. Harwood and I. a Guschina, *Biochimie*, 2009, **91**, 679–84.
64 A. Singh and S. I. Olsen, *Appl. Energy*, 2011, **88**, 3548–3555.
65 R. Subramaniam, S. Dufreche, M. Zappi and R. Bajpai, *J. Ind. Microbiol. Biotechnol.*, 2010, **37**, 1271–87.
66 X. Chen, Z. Li, X. Zhang, F. Hu, D. D. Y. Ryu and J. Bao, *Appl. Biochem. Biotechnol.*, 2009, **159**, 591–604.
67 X. Yu, Y. Zheng, K. M. Dorgan and S. Chen, *Bioresour. Technol.*, 2011, **102**, 6134–40.
68 Z. L. Liu, *Appl. Microbiol. Biotechnol.*, 2011, **90**, 809–25.
69 P. Meesters and G. Huijberts, *Appl. Microbiol.*, 1996, 575–579.
70 Y. Li, Z. (Kent) Zhao and F. Bai, *Enzyme Microb. Technol.*, 2007, **41**, 312–317.
71 S. Papanikolaou and G. Aggelis, *Eur. J. Lipid Sci. Technol.*, 2011, **113**, 1052–1073.
72 S. PAPANIKOLAOU, *Bioresour. Technol.*, 2004, **95**, 287–291.
73 I. Voss and A. Steinbüchel, *Appl. Microbiol. Biotechnol.*, 2001, **55**, 547–555.
74 A. Beopoulos, J.-M. Nicaud and C. Gaillardin, *Appl. Microbiol. Biotechnol.*, 2011, **90**, 1193–206.
75 M. Kosa and A. J. Ragauskas, *Trends Biotechnol.*, 2011, **29**, 53–61.
76 G. Christophe, V. Kumar, R. Nouaille, G. Gaudet, P. Fontanille, A. Pandey, C. R. Soccol and C. Larroche, *Brazilian Arch. Biol. Technol.*, 2012, **55**, 29–46.
77 H. G. Damude, H. Zhang, L. Farrall, K. G. Ripp, J.-F. Tomb, D. Hollerbach and N. S. Yadav, *Proc. Natl. Acad. Sci.*, 2006, **103**, 9446–9451.
78 D. Meesapyodsuk and X. Qiu, *Lipids*, 2012, **47**, 227–37.
79 J. G. Wallis, J. L. Watts and J. Browse, *Trends Biochem. Sci.*, 2002, **27**, 467–473.
80 F. Domergue, A. Abbadi, C. Ott, T. K. Zank, U. Zähringer and E. Heinz, *J. Biol. Chem.*, 2003, **278**, 35115–26.
81 H. Uemura, *Appl. Microbiol. Biotechnol.*, 2012, **95**, 1–12.
82 M. Li, X. Ou, X. Yang, D. Guo, X. Qian, L. Xing and M. Li, *Biotechnol. Lett.*, 2011, **33**, 1823–1830.
83 E. J. Brunner, P. J. S. Jones, S. Friel and M. Bartley, *Int. J. Epidemiol.*, 2008, **38**, 93–100.
84 I. Churruca, A. Fernández-Quintela and M. P. Portillo, *Biofactors*, **35**, 105–11.
85 J. M. Griinari, B. A. Corl, S. H. Lacy, P. Y. Chouinard, K. V Nurmela and D. E. Bauman, *J. Nutr.*, 2000, **130**, 2285–91.
86 H. Li, Y. Liu, Y. Bao, X. Liu and H. Zhang, *J. Food Sci.*, 2012, **77**, M330–6.
87 B. Zhang, C. Rong, H. Chen, Y. Song, H. Zhang and W. Chen, *Microb. Cell Fact.*, 2012, **11**, 51.
88 A. Liavonchanka, E. Hornung, I. Feussner and M. G. Rudolph, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 2576–81.
89 P. AKHTAR, J. I. GRAY and A. ASGHAR, *J. Food Lipids*, 1998, **5**, 283–297.
90 Y. Liang, T. Tang, T. Siddaramu, R. Choudhary and A. L. Umagiliyage, *Renew. Energy*, 2012, **40**, 130–136.
91 Y. Wang, Z. Gong, X. Yang, H. Shen, Q. Wang, J. Wang and Z. K. Zhao, *Process Biochem.*, 2015, **50**, 1097–1102.
92 G. Zhang, W. T. French, R. Hernandez, J. Hall, D. Sparks and W. E. Holmes, *J. Chem. Technol. Biotechnol.*, 2011, **86**, 642–650.
93 A. Ykema, E. C. Verbree, I. I. G. S. Verwoert, K. H. van der Linden, H. J. J. Nijkamp and H. Smit, *Appl. Microbiol. Biotechnol.*, 1990, **33**, 176–182.
94 M. Hassan, P. Blanc, L.-M. Granger, A. Pareilleux and G. Goma, *Appl. Microbiol. Biotechnol.*, 1993, **40**.

95 I. G. S. Verwoert, A. Ykema, J. C. Valkenburg, E. Verbree, H. John, J. Nijkamp and H. Smit, *Appl. Microbiol. Biotechnol.*, 1989, **32**, 327–333.
96 M. D. van de Rhee, P. M. Graça, H. J. Huizing and H. Mooibroek, *Mol. Gen. Genet.*, 1996, **250**, 252–8.
97 H. Rösel and G. Kunze, *Curr. Genet.*, 1998, **33**, 157–63.
98 K. Scholtmeijer, *Appl. Environ. Microbiol*, 2001, **67**.
99 M. Hanif, A. Pardo, M. Gorfer and M. Raudaskoski, *Curr. Genet.*, 2002, **41**, 183–188.
100 H. Yehuda, S. Droby, M. Wisniewski and M. Goldway, *Curr. Genet.*, 2001, **40**, 282–287.
101 R. J. N. Frandsen, J. a Andersson, M. B. Kristensen and H. Giese, *BMC Mol. Biol.*, 2008, **9**, 70.
102 Y. Liu, C. M. J. Koh, L. Sun, M. M. Hlaing, M. Du, N. Peng and L. Ji, *Appl. Microbiol. Biotechnol.*, 2013, **97**, 719–29.
103 L. Gritz and J. Davies, *Gene*, 1983, **25**, 179–188.
104 D. E. Brodersen, W. M. Clemons, A. P. Carter, R. J. Morgan-Warren, B. T. Wimberly and V. Ramakrishnan, *Cell*, 2000, **103**, 1143–1154.
105 N. Neuss, K. F. Koch, B. B. Molloy, W. Day, L. L. Huckstep, D. E. Dorman and J. D. Roberts, *Helv. Chim. Acta*, 1970, **53**, 2314–9.
106 C.-Y. Kuo, S.-Y. Chou and C.-T. Huang, *Appl. Microbiol. Biotechnol.*, 2004, **65**, 593–9.
107 T. Irie, Y. Honda, T. Hirano, T. Sato, H. Enei, T. Watanabe and M. Kuwahara, *Appl. Microbiol. Biotechnol.*, 2001, **56**, 707–709.
108 T. Sakaguchi, S. Amari, N. Nagashio, Y. Murakami, K. Yokoyama and E. Tamiya, *Biotechnol. Lett.*, 1998, **20**, 851–855.
109 a Domínguez, E. Fermiñán, M. Sánchez, F. J. González, F. M. Pérez-Campo, S. García, a B. Herrero, a San Vicente, J. Cabello, M. Prado, F. J. Iglesias, a Choupina, F. J. Burguillo, L. Fernández-Lago and M. C. López, *Int. Microbiol.*, 1998, **1**, 131–42.
110 A. Krivoruchko, V. Siewers and J. Nielsen, *Biotechnol. J.*, 2011, **6**, 262–76.
111 R.-B. Cheng, X.-Z. Lin, Z.-K. Wang, S.-J. Yang, H. Rong and Y. Ma, *World J. Microbiol. Biotechnol.*, 2010, **27**, 737–741.
112 G. Steinborn, G. Gellissen and G. Kunze, *FEMS Yeast Res.*, 2005, **5**, 1047–54.
113 E. Böer, M. Piontek and G. Kunze, *Appl. Microbiol. Biotechnol.*, 2009, **84**, 583–94.
114 R. D. Gietz and R. H. Schiestl, *Nat. Protoc.*, 2007, **2**, 31–35.
115 J. R. Thompson, E. Register, J. Curotto, M. Kurtz and R. Kelly, *Yeast*, 1998, **14**, 565–71.
116 G. Marchand, E. Fortier, B. Neveu, S. Bolduc, F. Belzile and R. R. Bélanger, *J. Microbiol. Methods*, 2007, **70**, 519–27.
117 C. M. McClelland, Y. C. Chang and K. J. Kwon-Chung, *Fungal Genet. Biol.*, 2005, **42**, 904–13.
118 R. J. N. Frandsen, *J. Microbiol. Methods*, 2011, **87**, 247–62.
119 C. B. Michielse, P. J. J. Hooykaas, C. a M. J. J. van den Hondel and A. F. J. Ram, *Curr. Genet.*, 2005, **48**, 1–17.
120 R. Hellens and P. Mullineaux, *Trends Plant Sci.*, 2000, **5**.
121 R. Kourist, F. Bracharz, J. Lorenzen, O. N. Kracht, M. Chovatia, C. Daum, S. Deshpande, A. Lipzen, M. Nolan, R. A. Ohm, I. V Grigoriev, S. Sun, J. Heitman, T. Brück and M. Nowrousian, *MBio*, 2015, **6**.
122 P. a Meesters, J. Springer and G. Eggink, *Appl. Microbiol. Biotechnol.*, 1997, **47**, 663–7.
123 A. Krogh, B. Larsson, G. von Heijne and E. L. Sonnhammer, *J. Mol. Biol.*, 2001, **305**, 567–80.
124 J. Blazeck, A. Hill, L. Liu, R. Knight, J. Miller, A. Pan, P. Otoupal and H. S. Alper, *Nat. Commun.*, 2014, **5**, 3131.
125 D. W. R. J. Sambrook, *Molecular cloning: a laboratory manual*, Cold Spring Harbor Laboratory, 2001.
126 P. Stothard, *Biotechniques*, 2000, **28**, 1102, 1104.
127 H. Chen, R. S. Nelson and J. L. Sherwood, *Biotechniques*, 1994, **16**, 664–8, 670.
128 J. FOLCH, M. LEES and G. H. SLOANE STANLEY, *J. Biol. Chem.*, 1957, **226**, 497–509.
129 M. J. Griffiths, R. P. van Hille and S. T. L. Harrison, *Lipids*, 2010, **45**, 1053–60.
130 J. C. Sacchettini and C. D. Poulter, *Science*, 1997, **277**, 1788–9.
131 K. T. Lane and L. S. Beese, *J. Lipid Res.*, 2006, **47**, 681–99.
132 J. Alonso-Gutierrez, R. Chan, T. S. Batth, P. D. Adams, J. D. Keasling, C. J. Petzold and T. S. Lee, *Metab. Eng.*, 2013, **19**, 33–41.
133 C. J. Paddon, P. J. Westfall, D. J. Pitera, K. Benjamin, K. Fisher, D. McPhee, M. D. Leavell, a Tai, a Main, D. Eng, D. R. Polichuk, K. H. Teoh, D. W. Reed, T. Treynor, J. Lenihan, M. Fleck, S. Bajad, G. Dang, D. Dengrove, D. Diola, G. Dorin, K. W. Ellens, S. Fickes, J. Galazzo, S. P. Gaucher, T. Geistlinger, R. Henry, M. Hepp, T. Horning, T. Iqbal, H. Jiang, L. Kizer, B. Lieu, D. Melis, N. Moss, R. Regentin, S. Secrest, H. Tsuruta, R. Vazquez, L. F. Westblade, L. Xu, M. Yu, Y. Zhang, L. Zhao, J. Lievense, P. S. Covello, J. D. Keasling, K. K. Reiling, N. S. Renninger and J. D. Newman, *Nature*, 2013, **496**, 528–32.
134 S. Malik, R. M. Cusidó, M. H. Mirjalili, E. Moyano, J. Palazón and M. Bonfill, *Process Biochem.*, 2011, **46**, 23–34.
135 V. Bartsch, *Pharm. Unserer Zeit*, 2005, **34**, 104–8.

136 H. P. Latscha, U. Kazmaier and H. A. Klein, *Organische Chemie*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.
137 C. A. SIMS, M. O. BALABAN and R. F. MAlTHEWS, *J. Food Sci.*, 1993, **58**, 1129–1131.
138 C.-H. Zhou, X. Xia, C.-X. Lin, D.-S. Tong and J. Beltramini, *Chem. Soc. Rev.*, 2011, **40**, 5588–617.
139 A. J. . Straathof, S. Panke and A. Schmid, *Curr. Opin. Biotechnol.*, 2002, **13**, 548–556.
140 L. B. Pickens, Y. Tang and Y.-H. Chooi, *Annu. Rev. Chem. Biomol. Eng.*, 2011, **2**, 211–36.
141 F. Alterthum and L. O. Ingram, *Appl. Environ. Microbiol.*, 1989, 55, 1943–1948.
142 H. W. Wisselink, M. J. Toirkens, Q. Wu, J. T. Pronk and A. J. A. van Maris, *Appl. Environ. Microbiol.*, 2009, **75**, 907–14.
143 L. Olsson and B. Hahn-Hägerdal, *Process Biochem.*, 1993, **28**, 249–257.
144 C. Khosla and J. D. Keasling, *Nat. Rev. Drug Discov.*, 2003, **2**, 1019–25.
145 J.-J. Zhong, *J. Biosci. Bioeng.*, 2002, **94**, 591–599.
146 P. K. Ajikumar, W.-H. Xiao, K. E. J. Tyo, Y. Wang, F. Simeon, E. Leonard, O. Mucha, T. H. Phon, B. Pfeifer and G. Stephanopoulos, *Science*, 2010, **330**, 70–4.
147 B. M. Lange, T. Rujan, W. Martin and R. Croteau, *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**, 13172–7.
148 T. KUZUYAMA, *Biosci. Biotechnol. Biochem.*, 2014, **66**, 1619–1627.
149 H. M. Miziorko, *Arch. Biochem. Biophys.*, 2011, **505**, 131–43.
150 M. Rohmer, *Nat. Prod. Rep.*, 1999, **16**, 565–574.
151 M. Rodríguez-Concepción and A. Boronat, *Plant Physiol.*, 2002, **130**, 1079–89.
152 M. Rodriguez-Concepcion, *Curr. Pharm. Des.*, 2004, **10**, 2391–2400.
153 S. C. Trapp and R. B. Croteau, *Genetics*, 2001, 158, 811–832.
154 P.-H. Liang, T.-P. Ko and A. H.-J. Wang, *Eur. J. Biochem.*, 2002, **269**, 3339–3354.
155 Y. Gao, R. B. Honzatko and R. J. Peters, *Nat. Prod. Rep.*, 2012, **29**, 1153–75.
156 C. A. Lesburg, J. M. Caruthers, C. M. Paschall and D. W. Christianson, *Curr. Opin. Struct. Biol.*, 1998, **8**, 695–703.
157 J. S. Dickschat, *Nat. Prod. Rep.*, 2011, **28**, 1917–36.
158 M. C. Y. Chang, R. a Eachus, W. Trieu, D.-K. Ro and J. D. Keasling, *Nat. Chem. Biol.*, 2007, **3**, 274–7.
159 R. Cao, Y. Zhang, F. M. Mann, C. Huang, D. Mukkamala, M. P. Hudock, M. E. Mead, S. Prisic, K. Wang, F.-Y. Lin, T.-K. Chang, R. J. Peters and E. Oldfield, *Proteins*, 2010, **78**, 2417–32.
160 T. Toyomasu, *Biosci. Biotechnol. Biochem.*, 2008, **72**, 1168–1175.
161 Y. Yamada, T. Kuzuyama, M. Komatsu, K. Shin-Ya, S. Omura, D. E. Cane and H. Ikeda, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, **112**, 857–62.
162 M. J. Smanski, R. M. Peterson, S.-X. Huang and B. Shen, *Curr. Opin. Chem. Biol.*, 2012, **16**, 132–41.
163 A. H. de Boer and I. J. de Vries-van Leeuwen, *Trends Plant Sci.*, 2012, **17**, 360–8.
164 R. Janke, C. Görner, M. Hirte, T. Brück and B. Loll, *Acta Crystallogr. D. Biol. Crystallogr.*, 2014, **70**, 1528–37.
165 H.-Y. Gong, Y. Zeng and X.-Y. Chen, *Nat. Products Bioprospect.*, 2014, **4**, 59–72.
166 P. Rabe and J. S. Dickschat, *Angew. Chem. Int. Ed. Engl.*, 2013, **52**, 1810–2.
167 M. Seemann, G. Zhai, J.-W. de Kraker, C. M. Paschall, D. W. Christianson and D. E. Cane, *J. Am. Chem. Soc.*, 2002, **124**, 7681–9.
168 B. Felicetti and D. E. Cane, *J. Am. Chem. Soc.*, 2004, **126**, 7212–21.
169 P. R. Wilderman and R. J. Peters, *J. Am. Chem. Soc.*, 2007, **129**, 15736–7.
170 D. Morrone, M. Xu, D. B. Fulton, M. K. Determan and R. J. Peters, *J. Am. Chem. Soc.*, 2008, **130**, 5400–1.
171 M. Xu, P. R. Wilderman and R. J. Peters, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 7397–401.
172 M. Köksal, H. Hu, R. M. Coates, R. J. Peters and D. W. Christianson, *Nat. Chem. Biol.*, 2011, **7**, 431–3.
173 B. Tudzynski, H. Kawaide and Y. Kamiya, *Curr. Genet.*, 1998, **34**, 234–240.
174 J. Söding, A. Biegert and A. N. Lupas, *Nucleic Acids Res.*, 2005, **33**, W244–8.
175 M. Köksal, Y. Jin, R. M. Coates, R. Croteau and D. W. Christianson, *Nature*, 2011, **469**, 116–20.
176 D. C. Hyatt, B. Youn, Y. Zhao, B. Santhamma, R. M. Coates, R. B. Croteau and C. Kang, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 5360–5.
177 L. S. Vedula, J. Jiang, T. Zakharian, D. E. Cane and D. W. Christianson, *Arch. Biochem. Biophys.*, 2008, **469**, 184–94.
178 E. Y. Shishova, L. Di Costanzo, D. E. Cane and D. W. Christianson, *Biochemistry*, 2007, **46**, 1941–51.
179 J. A. Aaron, X. Lin, D. E. Cane and D. W. Christianson, *Biochemistry*, 2010, **49**, 1787–97.
180 P. Baer, P. Rabe, K. Fischer, C. a Citron, T. a Klapschinski, M. Groll and J. S. Dickschat, *Angew. Chem. Int. Ed. Engl.*, 2014, **53**, 7652–6.
181 P. Schrepfer, A. Buettner, C. Goerner, M. Hertel, J. van Rijn, F. Wallrapp, W. Eisenreich, V. Sieber, R. Kourist and T. Brück, *Proc. Natl. Acad. Sci. U. S. A.*, 2016, **113**, E958–67.
182 O. D. Sparkman, *J. Am. Soc. Mass Spectrom.*, 2005, **16**, 1902–1903.
183 S. Bai and M. Jain, *Magn. Reson. Chem.*, 2008, **46**, 791–3.

184 I. Farkas and H. Pfander, *Helv. Chim. Acta*, 1990, **73**, 1980–1985.
185 A. Meguro, T. Tomita, M. Nishiyama and T. Kuzuyama, 2013.
186 R. Kazlauskas, P. Murphy, R. Wells, P. Schonholzer and J. Coll, *Aust. J. Chem.*, 1978, **31**, 1817.
187 M. Herin and B. Tursch, *Bull. des Sociétés Chim. Belges*, 2010, **85**, 707–719.
188 C. Ireland and D. J. Faulkner, *J. Org. Chem.*, 1977, **42**, 3157–3162.
189 A. Pardo-Vargas, F. a Ramos, C. C. Cirne-Santos, P. R. Stephens, I. C. P. Paixão, V. L. Teixeira and L. Castellanos, *Bioorg. Med. Chem. Lett.*, 2014, **24**, 4381–3.
190 A. Pardo-Vargas, I. de Barcelos Oliveira, P. R. S. Stephens, C. C. Cirne-Santos, I. C. N. de Palmer Paixão, F. A. Ramos, C. Jiménez, J. Rodríguez, J. A. L. C. Resende, V. L. Teixeira and L. Castellanos, *Mar. Drugs*, 2014, **12**, 4247–59.
191 X.-H. Cai, Y.-Y. Wang, P.-J. Zhao, Y. Li and X.-D. Luo, *Phytochemistry*, 2010, **71**, 1020–4.
192 C.-Y. Duh, M.-C. Chia, S.-K. Wang, H.-J. Chen, A. A. H. El-Gamal and C.-F. Dai, *J. Nat. Prod.*, 2001, **64**, 1028–1031.
193 J. Goodman, *J. Chem. Inf. Model.*, 2009, **49**, 2897–2898.
194 E. Ioannou, A. Quesada, M. M. Rahman, S. Gibbons, C. Vagias and V. Roussis, *European J. Org. Chem.*, 2012, **2012**, 5177–5186.
195 C. Tringali, M. Piattelli and G. Nicolosi, *Tetrahedron*, 1984, **40**, 799–803.
196 D. Morrone, L. Lowry, M. K. Determan, D. M. Hershey, M. Xu and R. J. Peters, *Appl. Microbiol. Biotechnol.*, 2010, **85**, 1893–906.
197 C. Görner, I. Häuslein, P. Schrepfer, W. Eisenreich and T. Brück, *ChemCatChem*, 2013, **5**, 3289–3298.
198 A. Meguro, Y. Motoyoshi, K. Teramoto, S. Ueda, Y. Totsuka, Y. Ando, T. Tomita, S.-Y. Kim, T. Kimura, M. Igarashi, R. Sawa, T. Shinada, M. Nishiyama and T. Kuzuyama, *Angew. Chemie*, 2015, **127**, 4427–4430.
199 M. J. Minch, *Concepts Magn. Reson.*, 1994, **6**, 41–56.
200 B. Yang, X.-F. Zhou, X.-P. Lin, J. Liu, Y. Peng, X.-W. Yang and Y. Liu, *Curr. Org. Chem.*, 2012, **16**, 1512–1539.
201 V. B. Urlacher and S. Eiben, *Trends Biotechnol.*, 2006, **24**, 324–30.
202 D. Werck-Reichhart and R. Feyereisen, *Genome Biol.*, 2000, **1**, REVIEWS3003.
203 B. Crešnar and S. Petrič, *Biochim. Biophys. Acta*, 2011, **1814**, 29–35.
204 V. B. Urlacher and M. Girhard, *Trends Biotechnol.*, 2012, **30**, 26–36.
205 E. G. Hrycay and S. M. Bandiera, *Arch. Biochem. Biophys.*, 2012, **522**, 71–89.
206 O. Shoji and Y. Watanabe, *J. Biol. Inorg. Chem.*, 2014, **19**, 529–39.
207 F. Hannemann, A. Bichet, K. M. Ewen and R. Bernhardt, *Biochim. Biophys. Acta*, 2007, **1770**, 330–44.
208 J. A. Peterson, M. C. Lorence and B. Amarneh, *J. Biol. Chem.*, 1990, **265**, 6066–73.
209 T. L. Poulos and E. R. Johnson, *Cytochrome P450*, Springer US, Boston, MA, 2005.
210 V. B. Urlacher, S. Lutz-Wahl and R. D. Schmid, *Appl. Microbiol. Biotechnol.*, 2004, **64**, 317–25.
211 J. a Peterson and S. E. Graham, *Structure*, 1998, **6**, 1079–1085.
212 I. F. Sevrioukova, C. Garcia, H. Li, B. Bhaskar and T. L. Poulos, *J. Mol. Biol.*, 2003, **333**, 377–392.
213 V. Y. Kuznetsov, T. L. Poulos and I. F. Sevrioukova, *Biochemistry*, 2006, **45**, 11934–44.
214 S. Tripathi, H. Li and T. L. Poulos, *Science*, 2013, **340**, 1227–30.
215 Y.-T. Lee, R. F. Wilson, I. Rupniewski and D. B. Goodin, *Biochemistry*, 2010, **49**, 3412–9.
216 R. A. Maplestone, M. J. Stone and D. H. Williams, *Gene*, 1992, **115**, 151–157.
217 J. Solecka, J. Zajko, M. Postek and A. Rajnisz, *Open Life Sci.*, 2012, **7**.
218 B. S. Deshpande, S. S. Ambedkar and J. G. Shewale, *Enzyme Microb. Technol.*, 1988, **10**, 455–473.
219 R. E. de L. Procópio, I. R. da Silva, M. K. Martins, J. L. de Azevedo and J. M. de Araújo, *Braz. J. Infect. Dis.*, **16**, 466–71.
220 C. Olano, F. Lombó, C. Méndez and J. A. Salas, *Metab. Eng.*, 2008, **10**, 281–92.
221 S. C. Wenzel and R. Müller, *Curr. Opin. Biotechnol.*, 2005, **16**, 594–606.
222 H. Zhang, Y. Wang, J. Wu, K. Skalina and B. A. Pfeifer, *Chem. Biol.*, 2010, **17**, 1232–40.
223 B. A. Pfeifer, S. J. Admiraal, H. Gramajo, D. E. Cane and C. Khosla, *Science*, 2001, **291**, 1790–2.
224 M. Jiang, G. Stephanopoulos and B. a Pfeifer, *Appl. Environ. Microbiol.*, 2012, **78**, 2497–504.
225 A. Celik, S. L. Flitsch and N. J. Turner, *Org. Biomol. Chem.*, 2005, **3**, 2930–4.
226 J. R. Anthony, L. C. Anthony, F. Nowroozi, G. Kwon, J. D. Newman and J. D. Keasling, *Metab. Eng.*, 2009, **11**, 13–9.
227 B. Zhao, X. Lin, L. Lei, D. C. Lamb, S. L. Kelly, M. R. Waterman and D. E. Cane, *J. Biol. Chem.*, 2008, **283**, 8183–9.
228 M. Girhard, T. Klaus, Y. Khatri, R. Bernhardt and V. B. Urlacher, *Appl. Microbiol. Biotechnol.*, 2010, **87**, 595–607.
229 W. Yang, S. G. Bell, H. Wang, W. Zhou, N. Hoskins, A. Dale, M. Bartlam, L.-L. Wong and Z. Rao, *J. Biol. Chem.*, 2010, **285**, 27372–84.
230 H. A. Hussain and J. M. Ward, *Appl. Environ. Microbiol.*, 2003, **69**, 373–82.

231 T. Makino, Y. Katsuyama, T. Otomatsu, N. Misawa and Y. Ohnishi, *Appl. Environ. Microbiol.*, 2014, **80**, 1371–9.

232 M. Xu, M. L. Hillwig, A. L. Lane, M. S. Tiernan, B. S. Moore and R. J. Peters, *J. Nat. Prod.*, 2014, **77**, 2144–7.

233 S. G. Bell, J. H. C. McMillan, J. A. Yorke, E. Kavanagh, E. O. D. Johnson and L.-L. Wong, *Chem. Commun. (Camb).*, 2012, **48**, 11692–4.

234 Y. Yasutake, T. Nishioka, N. Imoto and T. Tamura, *Chembiochem*, 2013, **14**, 2284–91.

235 P. Le-Huu, T. Heidt, B. Claasen, S. Laschat and V. B. Urlacher, *ACS Catal.*, 2015, **5**, 1772–1780.

236 D. E. Ward and C. K. Rhee, *Tetrahedron Lett.*, 1991, **32**, 7165–7166.

237 B. a Boghigian, M. Myint, J. Wu and B. a Pfeifer, *J. Ind. Microbiol. Biotechnol.*, 2011, **38**, 1809–20.

238 T. Aoyagi, T. Aoyama and F. Kojima, *J. Antibiot. (Tokyo).*, 1992, **45**, 1587–1591.

239 L. Crombie, G. Kneen, G. Pattenden and D. Whybrow, *J. Chem. Soc. Perkin Trans. 1*, 1980, 1711.

240 Y. Li, M. Carbone, R. M. Vitale, P. Amodeo, F. Castelluccio, G. Sicilia, E. Mollo, M. Nappo, G. Cimino, Y. Guo and M. Gavagnin, *J. Nat. Prod.*, 2010, **73**, 133–8.

241 N. KOMATSU, S. NAKAZAWA, M. HAMADA, M. SHIMO and T. TOMOSUGI, *J. Antibiot. (Tokyo).*, 1959, **12**, 12–6.

242 C. A. Citron, L. Barra, J. Wink and J. S. Dickschat, *Org. Biomol. Chem.*, 2015, **13**, 2673–83.

243 C. D. Costin, S. L. Hansen and D. P. Chambers, *J. Am. Oil Chem. Soc.*, 2008, **86**, 111–118.

244 H. Schäfer, K. Nau, A. Sickmann, R. Erdmann and H. E. Meyer, *Electrophoresis*, 2001, **22**, 2955–68.